NOTES ON

# Linear Algebra

Hao Su

Based on *Linear Algebra Done Right* (4th ed.).
Some exercises are omitted as trivial.

August 1, 2026

# Contents

CHAPTER 1

# Vector Spaces

## 1A $\mathbb{R}^n$ and $\mathbb{C}^n$

**Definition 1.1.** *Addition* and *multiplication* on $\mathbb{C}$ are defined by

$$(a + bi) + (c + di) = (a + c) + (b + d)i$$
$$(a + bi) \cdot (c + di) = (ac - bd) + (ad + bc)i$$

where $a, b, c, d \in \mathbb{R}$.

By properties of $\mathbb{R}$ and 1.1 we obtain properties of $\mathbb{C}$. By the existence of inverses we define $-\alpha$ and $\frac{1}{\alpha}$, and subtraction and division accordingly.

**Remark 1.2.** Use $\mathbb{F}$ (i.e., fields) to denote either $\mathbb{R}$ or $\mathbb{C}$. Elements of $\mathbb{F}$ are *scalars*. Say that $x_k$ is the $k^{\text{th}}$ *coordinate* of the *list* $(x_1, \ldots, x_n)$ ($n$, the *length* of the list, has to be finite). Lists, when thought of as arrows, are *vectors*. Addition and scalar multiplication on lists are defined componentwise in the standard way.

### Exercises

**Exercise 5.** Additive inverse of complex arithmetic.

This is due to the additive inverse of real arithmetic.

**Exercise 6.** Multiplicative inverse of complex arithmetic.

For $\alpha = a + bi$, where $a \neq 0$, we have $\frac{a - bi}{a^2 + b^2}$ as the multiplicative inverse of $\alpha$. Thus one exists. Also all the multiplicative inverses of $\alpha$ equals it by real arithmetic.

**Exercise 8.** Find two distinct square roots of $i$.

We solve the equation $(a + bi)^2 = i$ to obtain $\pm\left(\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i\right)$.

## 1B Definition of Vector Space

We need a space where things act like vectors, which is defined as follows.

**Definition 1.3.** A *vector space* over a field $\mathbb{F}$ is a set $V$ equipped with two operations: vector addition $+ : V \times V \to V$, and scalar multiplication $\cdot : \mathbb{F} \times V \to V$, satisfying the following axioms:

1. $(V, +)$ is an abelian group, i.e.:
   (a) (Associativity) $u + (v + w) = (u + v) + w$ for all $u, v, w \in V$,
   (b) (Commutativity) $u + v = v + u$ for all $u, v \in V$,
   (c) (Identity) There exists an element $0 \in V$ such that $v + 0 = v$ for all $v \in V$,
   (d) (Inverses) For each $v \in V$, there exists an element $-v \in V$ such that $v + (-v) = 0$.
2. Scalar multiplication satisfies:
   (a) (Multiplicative identity) $1 \cdot v = v$ for all $v \in V$,
   (b) (Associativity) $a \cdot (b \cdot v) = (ab) \cdot v$ for all $a, b \in \mathbb{F}, v \in V$,
   (c) (Distributivity over vector addition) $a \cdot (u + v) = a \cdot u + a \cdot v$ for all $a \in \mathbb{F}$, $u, v \in V$,
   (d) (Distributivity over field addition) $(a + b) \cdot v = a \cdot v + b \cdot v$ for all $a, b \in \mathbb{F}, v \in V$.

The simplest vector space is $\{0\}$.

**Definition 1.4.** Elements of a vector space are called *vectors* or *points*.

**Example 1.5.** For $f, g \in \mathbb{F}^S$ and $\lambda \in \mathbb{F}$, $f + g \in \mathbb{F}^S$ is defined by $\forall x \in S, (f + g)(x) = f(x) + g(x)$ and $\lambda f \in \mathbb{F}^S$ by $\forall x (\lambda f)(x) = \lambda f(x)$. If $S \neq \emptyset$, then $\mathbb{F}^S$ is a vector space over $\mathbb{F}$. The vector space $\mathbb{F}^n$ is a special case of $\mathbb{F}^S$, where $S = \{1, 2, \dots, n\}$.

**Theorem 1.6.** *A vector space has a unique additive identity.*

*Proof.* Suppose $0$ and $0'$ are both additive identities for some vector space $V$. Then

$$0' = 0' + 0 = 0 + 0' = 0.$$ ⊣

**Theorem 1.7.** *Every element in a vector space has a unique additive inverse.*

*Proof.* Suppose $a \in V$ has two additive inverses $b$ and $c$. Then

$$b = b + (a + c) = (b + a) + c = c.$$ ⊣

Notations like $-v$ and $w - v$ make sense due to the uniqueness of additive inverses. From now on, $V$ denotes a vector space over $\mathbb{F}$.

**Theorem 1.8.** $\forall v \in V, 0v = 0$.

*Proof.* We have for any $v \in V$

$$0v = (0 + 0)v = 0v + 0v.$$

Adding the additive inverse of $0v$ to both sides of the equation above gives $0 = 0v$. ⊣

*Comment.* We *have* to use distributivity for that's where vector addition and scalar multiplication are connected in 1.3. The first equation holds because $0 \in \mathbb{F}$.

**Theorem 1.9.** $\forall a \in \mathbb{F}, a0 = 0$.

*Proof.* We have for any $a \in \mathbb{F}$

$$a0 = a(0 + 0) = a0 + a0.$$

Adding the additive inverse of $a0$ to both sides of the equation above gives $0 = a0$. ⊣

Similarly, $0 = (1 + (-1))v$ gives us

**Theorem 1.10.** $\forall v \in V, (-1)v = -v$.

## Exercises

**Exercise 1.** Prove that $\forall v \in V, -(-v) = v$.

$0 = (-v) + (-(-v)) = v + (-v)$.

**Exercise 2.** Suppose $a \in \mathbb{F}, v \in V$, and $av = 0$. Prove that $a = 0$ or $v = 0$.

If $a \neq 0$ we have that $v = (a \cdot \frac{1}{a})v = \frac{1}{a} \cdot (av) = 0$, which is equivalent to what we are asked to prove.

**Exercise 5.** Show that (d) of item 1 in 1.3 can be replaced with 1.8.

We are to show the existence of additive inverse from 1.8 and the rest of 1.3.

$$0 = 0v = (1 - 1)v = v + (-1)v.$$

Thus the additive inverse of $v$ exists, namely $(-1)v$.

**Exercise 6.** Let $\infty$ and $-\infty$ denote two distinct objects, neither of which is in $\mathbb{R}$. Define an addition and scalar multiplication on $\mathbb{R} \cup \{\infty, -\infty\}$: the sum and product of two real numbers is as usual, and for $t \in \mathbb{R}$ define

$$t\infty = \begin{cases} -\infty & \text{if } t < 0, \\ 0 & \text{if } t = 0, \\ \infty & \text{if } t > 0, \end{cases} \quad t(-\infty) = \begin{cases} \infty & \text{if } t < 0, \\ 0 & \text{if } t = 0, \\ -\infty & \text{if } t > 0, \end{cases}$$

and

$$\begin{aligned} t + \infty = \infty + t = \infty + \infty = \infty, \\ t + (-\infty) = (-\infty) + t = (-\infty) + (-\infty) = -\infty, \\ \infty + (-\infty) = (-\infty) + \infty = 0. \end{aligned}$$

With these operations of addition and scalar multiplication, is $\mathbb{R} \cup \{\infty, -\infty\}$ a vector space over $\mathbb{R}$? Explain.

Consider $(\mathbb{R} \cup \{\infty, -\infty\}, +)$. Commutativity and existence of identity and inverses holds by definition. However $(u, v, w) = (3, \infty, -\infty)$ violates commutativity, hence $(\mathbb{R} \cup \{\infty, -\infty\}, +)$ is not an abelian group, thus $\mathbb{R} \cup \{\infty, -\infty\}$ not a vector space.

**Exercise 8.** Suppose $V$ is a real vector space.

- The *complexification* of $V$, denoted by $V_{\mathbb{C}}$, equals $V \times V$. An element of $V_{\mathbb{C}}$ is an ordered pair $(u, v)$, where $u, v \in V$, but we write this as $u + iv$.
- Addition on $V_{\mathbb{C}}$ is defined by

$$(u_1 + iv_1) + (u_2 + iv_2) = (u_1 + u_2) + i(v_1 + v_2)$$

  for all $u_1, v_1, u_2, v_2 \in V$.
- Complex scalar multiplication on $V_{\mathbb{C}}$ is defined by

$$(a + bi)(u + iv) = (au - bv) + i(av + bu)$$

  for all $a, b \in \mathbb{R}$ and all $u, v \in V$.

Prove that with the definitions of addition and scalar multiplication as above, $V_{\mathbb{C}}$ is a complex vector space.

We verify each of the requirements specified by 1.3 by properties of $V$ as an vector space, very much like verifying the properties of $\mathbb{C}$ by those of $\mathbb{R}$.

*Comment.* Think of $V$ as a subset of $V_{\mathbb{C}}$ by identifying $u \in V$ with $u + i0$. The construction of $V_{\mathbb{C}}$ can then be thought of as generalizing the construction of $\mathbb{C}^n$ from $\mathbb{R}^n$ (thought of as a subset of $\mathbb{C}^n$.)

## 1C Subspaces

**Definition 1.11.** A subset $U$ of $V$ is called a *subspace* of $V$ if $U$ is also a vector space with the same additive identity, addition, and scalar multiplication as on $V$.

**Theorem 1.12.** *$U \subseteq V$ is a subspace of $V$ iff it (1) has the additive identity of $V$ (or is nonempty, because we can take $u \in U$ then $0u \in U$.), (2) is closed under addtion, and (3) is closed under scalar multiplication.*

*Proof.* Both directions hold by definition. In paticular, closure ensures that addition and multiplication are reasonably defined ($U \times U \to U$ and $\mathbb{F} \times U \to U$) and properties such as associativity hold because they hold on $V$ and $U \subseteq V$. ⊣

**Example 1.13.** The set of differentiable real-valued functions $f$ on the interval (0,3) such that $f'(2) = b$ is a subspace of $\mathbb{R}^{(0,3)}$ iff $b = 0$ for closure under scalar multiplication, which shows the linear structure underlying parts of calculus. The subspaces of $\mathbb{R}^2$ are precisely $\{0\}$ all lines in $\mathbb{R}^2$ containing the origin and $\mathbb{R}$, which intuitively justifies the word “linear”.

**Definition 1.14.** Suppose $V_1, \dots, V_m$ are subspaces of $V$. The *sum* of them is

$$V_1 + \cdots + V_m = \{v_1 + \cdots + v_m \mid \bigwedge v_i \in V_i\}$$

**Theorem 1.15.** *Suppose $V_1, \dots, V_m$ are subspaces of $V$. Then $V_1 + \cdots + V_m$ is the smallest subspace of $V$ containing $V_1, \dots, V_m$.*

*Proof.* That $V_1 + \cdots + V_m$ is a subspace and contains $V_1, \dots, V_m$ is trivial. Suppose that $V'$ contains $V_1, \dots, V_m$ and is a subspace. By 1.14 and closure under addition we have that $V_1 + \cdots + V_m \subseteq V'$, thus the minimality. ⊣

**Definition 1.16.** Suppose $V_1, \dots, V_m$ are subspaces of $V$. The sum $V_1 + \cdots + V_m$ is called a *direct sum* if each element of $V_1 + \cdots + V_m$ can be written in only one way as a sum $v_1 + \cdots + v_m$ where each $v_k \in V_k$, denoted $V_1 \oplus \cdots \oplus V_m$.

The definition of direct sum requires every vector in the sum to have a unique representation as an appropriate sum.

**Theorem 1.17.** *Suppose $V_1, \dots, V_m$ are subspaces of $V$. Then $V_1 + \cdots + V_m$ is a direct sum iff the only way to write 0 as a sum $v_1 + \cdots + v_m$, where each $v_k \in V_k$, is by taking each $v_k$ equal to 0.*

*Proof.* ($\Rightarrow$ is trivial. Consider $\Leftarrow$.) Suppose for sake of contradiction that $V_1 + \cdots + V_m$ is *not* a direct sum. Then there exists $v \in V_1 + \cdots + V_m$ such that $v = v_1' + \cdots + v_m' = v_1'' + \cdots + v_m''$, where $v_k' \in V_k$ and $v_k'' \in V_k$ for each $k$ and $(v_1', \dots, v_m') \neq (v_1'', \dots, v_m'')$. Then we have $0 = (v_1' - v_1'') + \cdots + (v_m' - v_m'')$ and at least a $j$ such that $v_j' - v_j'' \neq 0$. Hence contradiction. ⊣

**Theorem 1.18.** *Suppose $U$ and $W$ are subspaces of $V$. Then*

$$U + W \text{ is a direct sum} \Leftrightarrow U \cap W = \{0\}.$$

*Proof.* $\Rightarrow$: Say that $v \in U \cap W$ and $v \neq 0$, then $0 = v + (-v)$, where $v \in U$ and $-v \in W$, hence $U + W$ is not a direct sum, and contradiction.
$\Leftarrow$: Say that $0 = a + b$, where $a \in U$, $b \in W$, and $a \neq 0$. Then $b = -a \in U$. Hence $b \in U \cap W$ and $b \neq 0$, and contradiction. $\dashv$

Sums of subspaces are analogous to unions of subsets. Similarly, direct sums of subspaces are analogous to disjoint unions of subsets.

## Exercises

**Exercise 6.** Is $\{(a, b, c) \in \mathbb{R}^3 \mid a^3 = b^3\}$ a subspace of $\mathbb{R}^3$? How about $\mathbb{C}^3$?

By $a^3 = b^3$ we have that $(a - b)(a^2 + ab + b^2) = 0$. That is, if $b \neq 0$ then $\frac{a}{b} \in \left\{1, \frac{-1 \pm \sqrt{3}i}{2}\right\}$. We need a unique $\frac{a}{b}$ so that it is closed under addition, which holds in $\mathbb{R}$ but not in $\mathbb{C}$.

**Exercise 7.** Prove or give a counterexample: If $U$ is a nonempty subset of $\mathbb{R}^2$ such that $U$ is closed under addition and contains additive inverses, then $U$ is a subspace of $\mathbb{R}^2$.

A counterexample is $\mathbb{Z}^2$.

**Exercise 8.** Give an example of a nonempty subset $U$ of $\mathbb{R}^2$ such that $U$ is closed under scalar multiplication, but $U$ is not a subspace of $\mathbb{R}$.

Cf. 1C.6. Easily we have that $\{(a, b) \mid a = 2b \vee a = b\}$ is not closed under addition. Note that $\{(a, b) \mid a = 2b\} + \{(a, b) \mid a = b\}$ is a direct sum, yet their union is not a subspace.

**Exercise 12.** Prove that the union of two subspaces of $V$ is a subspace of $V$ iff one of the subspaces is a subset of another.

Suppose otherwise that $U$ and $W$ are subspaces of $V$, $u \in U \setminus W$, and $w \in W \setminus U$. Then $u + w \notin U \cup W$, for that otherwise contradiction will arise. Hence $U \cup W$ is not a subspace of $V$.

**Exercise 13.** Prove that the union of three subspaces of $V$ is a subspace of $V$ iff one of the subspaces contains the other two.

to do

> **Exercise 18.** Does the operation of addition on the subspaces of $V$ have an additive identity? Which subspaces have additive inverses?

$\{0\}$ is *the* additive identity of addition on subspaces of $V$. Also only $\{0\}$ has an additive inverse, namely itself. Suppose otherwise for sake of contradiction that $W \subseteq V$ has an additive inverse $-W$, $w \in W$, and $w \neq 0$. Then $w = w + 0 \in W + (-W) = \{0\}$, hence contradiction.

> **Exercise 19.** Prove or give a counterexample: If $V_1, V_2, U$ are subspaces of $V$ such that $V_1 + U = V_2 + U$ then $V_1 = V_2$.

$V_1 + U = V_2 + U$ should always hold, as long as $V_1 \subseteq U$ and $V_2 \subseteq U$, considering 1.15.

> **Exercise 20.** Suppose $U = \{(x, x, y, y) \in \mathbb{F}^4 \mid x, y \in \mathbb{F}\}$. Find a subspace $W$ of $\mathbb{F}^4$ such that $\mathbb{F}^4 = U \oplus W$.

It is easy to verify that $\{(x, -x, y, -y) \in \mathbb{F}^4 \mid x, y \in \mathbb{F}\}$ suffices.

> **Exercise 22.** Suppose
>
> $$U = \{(x, y, x + y, x - y, 2x) \in \mathbb{F}^5 \ : \ x, y \in \mathbb{F}\}.$$
>
> Find three subspaces $W_1$, $W_2$, $W_3$ of $\mathbb{F}^5$, none of which equals $\{0\}$, such that
>
> $$\mathbb{F}^5 = U \oplus W_1 \oplus W_2 \oplus W_3.$$

We may follow the instinct that $(x, y, x + y + a, x - y + b, 2x + c)$ 'discribes' $\mathbb{F}^5$, which gives $(0, 0, 1, 0, 0), (0, 0, 0, 1, 0)$, and $(0, 0, 0, 0, 1)$. And fortunately, it follows that these spaces indeed make a direct sum.

> **Exercise 23.** Prove or give a counterexample: If $V_1, V_2, U$ are subspaces of $V$ such that $V = V_1 \oplus U = V_2 \oplus U$ then $V_1 = V_2$.

A counterexample is that $V_1 = \{(x, 0) \mid x \in \mathbb{R}\}, V_2 = \{(x, x) \mid x \in \mathbb{R}\}, U = \{(0, y) \mid y \in \mathbb{R}\}$, and $V = \mathbb{R}^2$. The intuition is well illustrated by the counterexample: there are different sets of vectors that span a plane, which may share same components.

**Exercise 24.** Let $V_e$ denote the set of real-valued even functions on $\mathbb{R}$ and $V_o$ the set of real-valued odd functions. Show that $\mathbb{R}^{\mathbb{R}} = V_e \oplus V_o$.

Trivial considering that $f(x) = \frac{1}{2}(f(x) + f(-x)) + \frac{1}{2}(f(x) - f(-x))$.

Chapter 2

# Finite-Dimensional Vector Spaces

## 2A Span and Linear Independence

We will usually write lists of vectors without surrounding parentheses.

**Definition 2.1.** A *linear combination* of a list $v_1, \ldots, v_m$ of vectors in $V$ is a vector of the form

$$a_1 v_1 + \cdots + a_m v_m,$$

where $a_1, \ldots, a_m \in \mathbb{F}$.

**Definition 2.2.** The set of all linear combinations of a list of vectors $v_1, \ldots, v_m$ in $V$ is called the *span* of $v_1, \ldots, v_m$, denoted $\operatorname{span}(v_1, \ldots, v_m)$. The span of the empty list $()$ is defined to be $\{0\}$.

**Theorem 2.3.** *The span of a list of vectors in $V$ is the smallest subspace of $V$ containing all vectors in the list.*

The proof is omitted as trivial.

**Definition 2.4.** If $\operatorname{span}(v_1, \ldots, v_m)$ equals $V$, say that the list $v_1, \ldots, v_m$ *spans* $V$.

**Definition 2.5.** A vector space is called *finite-dimensional* if some list of vectors in it spans the space. A vector space is called *infinite-dimensional* if it is not finite-dimensional.

**Definition 2.6.** A function $p : \mathbb{F} \to \mathbb{F}$ is called a *polynomial* with coefficients in $\mathbb{F}$ if there exist $a_0, \dots, a_m \in \mathbb{F}$ such that

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m$$

for all $z \in \mathbb{F}$. $\mathcal{P}(\mathbb{F})$ is the set of all polynomials with coefficients in $\mathbb{F}$.

4.5 justifies the following definition.

**Definition 2.7.** A polynomial $p \in \mathcal{P}(\mathbb{F})$ is said to have *degree* $m$ if there exist scalars $a_0, a_1, \dots, a_m \in \mathbb{F}$ with $a_m \neq 0$ such that for every $z \in \mathbb{F}$, we have

$$p(z) = a_0 + a_1 z + \cdots + a_m z^m.$$

The polynomial that is identically 0 is said to have degree $-\infty$. The degree of a polynomial $p$ is denoted by $\deg p$. For $m$ a nonnegative integer, $\mathcal{P}_m(\mathbb{F})$ denotes the set of all polynomials with coefficients in $\mathbb{F}$ and degree at most $m$.

$\mathcal{P}_m(\mathbb{F}) = \operatorname{span}(1, z, \dots, z^m)$ and thus a finite-dimensional vector space for each $m \in \mathbb{N}$. Consider any list of elements of $\mathcal{P}(\mathbb{F})$. Let $m$ denote the highest degree of the polynomials in this list. Then every polynomial in the span of this list has degree at most $m$. Thus $z^{m+1}$ is not in the span of our list. Hence no list spans $\mathcal{P}(\mathbb{F})$. Thus $\mathcal{P}(\mathbb{F})$ is infinite-dimensional.

**Definition 2.8.** A list $v_1, \dots, v_m$ of vectors in $V$ is called *linearly independent* if the only choice of $a_1, \dots, a_m \in \mathbb{F}$ that makes

$$a_1 v_1 + \cdots + a_m v_m = 0$$

is $a_1 = \cdots = a_m = 0$. The empty list () is declared to be linearly independent.

**Remark 2.9.** $\operatorname{span}(v_1, \dots, v_m)$ is by definition $\operatorname{span}(v_1) + \cdots + \operatorname{span}(v_m)$ and is a direct sum iff $v_1, \dots, v_m$ is linearly independent. Cf. 1.17.

If some vectors are removed from a linearly independent list, the remaining list is also linearly independent.

**Definition 2.10.** A list of vectors in $V$ is called *linearly dependent* if it is not linearly independent.

If some vector in a list of vectors in $V$ is a linear combination of the other vectors, then the list is linearly dependent.

**Theorem 2.11.** *Suppose $v_1, \ldots, v_m$ is a linearly dependent list in $V$. Then there exists $k \in \{1, 2, \ldots, m\}$ such that*

$$v_k \in \operatorname{span}(v_1, \ldots, v_{k-1}).$$

*Furthermore, if $k$ satisfies the condition above and the $k^{th}$ term is removed from $v_1, \ldots, v_m$, then the span of the remaining list equals* $\operatorname{span}(v_1, \ldots, v_m)$.

*Proof.* We prove by induction the contraposition of this theorem.

Base case: $m = 1$, if $v_1 \notin \operatorname{span}()$, then $V$ is liearly independent.

Inductive case: suppose that for $m = j$, if

$$\forall k \in \{1, 2, \ldots, m\}, v_k \notin \operatorname{span}(v_1, \ldots, v_{k-1}),$$

then $V$ is linearly independent. Now consider the case where $m = j + 1$. By IH we have that $(v_1, \ldots, v_{m-1})$ is linearly independent. Suppose, however, that $(v_1, \ldots, v_m)$ is not. Then $0 = a_1v_1 + \cdots + a_mv_m$, where $(a_1, \ldots, a_m) \neq 0$. Suppose that $a_m = 0$, then we have that $0 = a_1v_1 + \cdots + a_{m-1}v_{m-1}$, where $(a_1, \ldots, a_{m-1}) \neq 0$, contradicting that $(v_1, \ldots, v_{m-1})$ is linearly independent.Thus $a_m \neq 0$ and $v_m \in \operatorname{span}(v_1, \ldots, v_{m-1})$, contradicting the assumption. Hence $(v_1, \ldots, v_m)$ is linearly independent.

The proof of the furthermore part of the theorem is ommitted as trivial. ⊣

**Theorem 2.12.** *In a finite-dimensional vector space, the length of every linearly independent list of vectors is less than or equal to the length of every spanning list of vectors.*

*Proof.* Suppose that $u_1, \ldots, u_m$ is linearly independent in $V$ and let $B$ be the list $w_1, \ldots, w_n$ spans $V$. We adjoin one by one $u_i$ at the beginning of $B$, placing it after $u_1, \ldots, u_{i-1}$ and remove one of the $w$'s each time a $u$ is adjoined according to 2.11. No $u$'s are removed in this process, and hence there are at least as many $w$'s as $u$'s. ⊣

**Theorem 2.13.** *Subspaces of a finite-dimensional vector space are finite-dimensional.*

*Proof.* Say that $V$ is finite dimensional and $U$ a subspace of it. Do the following until a list of $u$'s that spans $U$ is constructed: pick $u_i$ in $U$ such that

$$u_i \notin \operatorname{span}(u_1, \ldots, u_{i-1})$$

By 2.11 the list constructed is linearly independent. By 2.12 the list of $u$'s is finite, and the process terminates. ⊣

## Exercises

**Exercise 3.** Suppose $v_1, \dots, v_m$ is a list of vectors in $V$. For $k \in \{1, \dots, m\}$, let

$$w_k = v_1 + \dots + v_k.$$

Show that $\operatorname{span}(v_1, \dots, v_m) = \operatorname{span}(w_1, \dots, w_m)$.

We solve the equation

$$a_1 v_1 + \dots + a_m v_m = b_1 w_1 + \dots + b_m w_m$$

to obtain

$$(b_1, \dots, b_{m-1}, b_m) = (a_1 - a_2, \dots, a_{m-1} - a_m, a_m).$$

**Exercise 11.** Prove or give a counterexample: If $v_1, \dots, v_m$ and $w_1, \dots, w_m$ are linearly independent lists of vectors in $V$, then the list $v_1 + w_1, \dots, v_m + w_m$ is linearly independent.

Counterexample: $v_1 = 1$, $w_1 = -1$.

**Exercise 17.** Prove that $V$ is infinite-dimensional if and only if there is a sequence $v_1, v_2, \dots$ of vectors in $V$ such that $v_1, \dots, v_m$ is linearly independent for every positive integer $m$.

$\Leftarrow$: Say that $V$ is finite-dimensional. Then a list of length $m$ spans it. Yet we can find a linearly independent list of length $m + 1$, contradicting 2.12.
$\Rightarrow$: We can construct a sequence adjoining vectors that are not in the span of the vectors already included in the sequence. $V$ is infinite-dimensional, so no list spans $V$, thus we can always find some vector to adjoin. Hence the sequence constructed is indeed infinite.

**Exercise 18.** Prove that $\mathbb{F}^\infty$ is infinite-dimensional.

Consider the sequence $(1, \dots), (0, 1, \dots), (0, 0, 1, \dots), \dots$.

**Exercise 19.** Prove that the real vector space of all continuous real-valued functions on the interval $[0, 1]$ is infinite-dimensional.

Consider the sequence $1, x, x^2, \dots$.

**Exercise 20.** Suppose $p_0, p_1, \dots, p_m$ are polynomials in $\mathcal{P}_m(\mathbb{F})$ such that $p_k(2) = 0$ for each $k \in \{0, \dots, m\}$. Prove that $p_0, p_1, \dots, p_m$ is not linearly independent in $\mathcal{P}_m(\mathbb{F})$.

(Cf. 2.24. We can avoid dimension here.) Say that it is. Therefore it spans $P_m(\mathbb{F})$ (for that if not, we can find a linearly independent list of length $(m+2)$, contradicting a spanning list $1, x, \dots, x_m$ by 2.11.) Yet there exist polynomials that are not in its span, hence contradiction.

## 2B Bases

**Definition 2.14.** A *basis* of $V$ is a list of vectors in $V$ that is linearly independent and spans $V$.

**Theorem 2.15.** *A list $v_1, \dots, v_n$ of vectors in $V$ is a basis of $V$ iff every $v \in V$ can be written uniquely in the form*

$$v = a_1 v_1 + \cdots + a_n v_n,$$

*where $a_1, \dots, a_n \in \mathbb{F}$.*

Both directions are trivial.

**Theorem 2.16.** *Every spanning list in a vector space can be reduced to a basis of the vector space.*

*Proof.* We remove the first $v_k$ each time until there is no $v$ to remove accroding to 2.11. ⊣

**Corollary 2.17.** *Every finite-dimensional vector space has a basis.*

**Theorem 2.18.** *Every linearly independent list in a finite-dimensional vector space extends to a basis.*

*Proof.* Suppose $u_1, \dots, u_m$ is linearly independent in a finite-dimensional vector space $V$. Say that it is not a basis of $V$. Start with it and adjoin one by one

$$w_i \notin \operatorname{span}(u_1, \dots, u_m, w_1, \dots, w_{i-1})$$

where $w_0 = u_m$ and $i \in \mathbb{Z}_+$. We finally will arive at a basis of $V$. Cf. 2.13. ⊣

We may alternatively append a spanning list of $V$ to $u_1, \dots, u_m$ and reduce the list we get to a basis, cf. 2.16.

**Theorem 2.19.** *Suppose $V$ is finite-dimensional and $U$ is a subspace of $V$. Then there is a subspace $W$ of $V$ such that $V = U \oplus W$.*

This is trivial, for that the added $w$'s in 2.18 spans $W$. Also cf. 2.9.

## Exercises

**Exercise 5.** Suppose $V$ is finite-dimensional and $U, W$ are subspaces of $V$ such that $V = U + W$. Prove that there exists a basis of $V$ consisting of vectors in $U \cup W$.

We take the bases of $U$ and $W$, adjoin them and reduce the list to a linearly independent list.

**Exercise 8.** Prove or give a counterexample: If $v_1, v_2, v_3, v_4$ is a basis of $V$ and $U$ is a subspace of $V$ such that $v_1, v_2 \in U$ and $v_3 \notin U$ and $v_4 \notin U$, then $v_1, v_2$ is a basis of $U$.

Counterexample: $v_1 = (1, 0, 0, 0), v_2 = (0, 1, 0, 0), v_3 = (0, 0, 1, 0), v_4 = (0, 0, 0, 1)$.
$U = \operatorname{span}((1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 1))$.

**Exercise 11.** Suppose $V$ is a real vector space. Show that if $v_1, \dots, v_n$ is a basis of $V$ (as a real vector space), then $v_1, \dots, v_n$ is also a basis of the complexification $V_{\mathbb{C}}$ (as a complex vector space). Cf. 1B.8.

$v_k$ is thought of as $v_k + 0u_k i$ in $V_{\mathbb{C}}$. To show that $v_1, \dots, v_n$ is linearly independent, let

$$0 = \sum_{j=1}^{n} (a_j + b_j i) v_j,$$

which gives

$$\sum_{j=1}^{n} a_j v_j = \sum_{j=1}^{n} b_j v_j = 0,$$

which in turn shows the linear independence given that $v_1, \dots, v_n$ are linearly independent in $V$. That $v_1, \dots, v_n$ spans $V_{\mathbb{C}}$ is due to each element in $V_{\mathbb{C}}$ can be written $a + bi$, where $a, b \in \operatorname{span}(v_1, \dots, v_n)$.

## 2C Dimension

**Theorem 2.20.** *Any two bases of a finite-dimensional vector space have the same length.*

This is trivial, given 2.12.

**Definition 2.21.** The *dimension* of a finite-dimensional vector space is the length of any basis of the vector space, denoted by $\dim V$.

**Theorem 2.22.** *If $V$ is finite-dimensional and $U$ is a subspace of $V$, then $\dim U \leq \dim V$.*

This statement follows from 2.18. Also, the basis of $V$ is a spanning list and the basis of $U$ a linearly independent one, thus 2.11 also implies this.

**Remark 2.23.** The choice of $\mathbb{F}$ matters in identifying the dimension of a vector space. Verify that $\mathbb{C}$ has dimension one over $\mathbb{C}$ and two over $\mathbb{R}$.

**Theorem 2.24.** *Suppose $V$ is finite-dimensional. Then every linearly independent list of vectors in $V$ of length $\dim V$ is a basis of $V$.*

*Proof.* Say that a linearly independent list $v_1, \ldots, v_m$, where $m = \dim V$, does not span $V$, and $w \notin \operatorname{span}(v_1, \ldots, v_m)$. Then $v_1, \ldots, v_m, w$ forms a linealy independent list of length $\dim V + 1$, contradicting 2.11. ⊣

**Corollary 2.25.** *Suppose that $V$ is finite-dimensional and $U$ is a subspace of $V$ such that $\dim U = \dim V$. Then $U = V$.*

**Theorem 2.26.** *Suppose $V$ is finite-dimensional. Then every spanning list of vectors in $V$ of length $\dim V$ is a basis of $V$.*

*Proof.* Say that a spanning list of $V$, $v_1, \ldots, v_m$ is not a linearly independent one. Then we can remove $v_k$ such that $v_k \in \operatorname{span}(v_1, \ldots, v_{k-1})$, contradicting 2.11. ⊣

**Theorem 2.27.** *If $V_1$ and $V_2$ are subspaces of a finite-dimensional vector space, then*

$$\dim(V_1 + V_2) = \dim V_1 + \dim V_2 - \dim(V_1 \cap V_2).$$

*Proof.* We extend the basis of $V_1 \cap V_2$ to bases of $V_1$ and $V_2$, obtaining three bases, and show that they are linearly independent.

Say that $e_1, \dots, e_m$ is the basis of $V_1 \cap V_2$. By 2.18 and 2.19 we can extend it to the bases of $V_1$ and $V_2$ by adding $f_1, \dots, f_n$ and $g_1, \dots, g_k$ respectively, where $e_1, \dots, e_m, f_1, \dots, f_n$ and $e_1, \dots, e_m, g_1, \dots, g_k$ are both linearly independent. Now obviously $V_1 + V_2 = \operatorname{span}(e_1, \dots, e_m, f_1, \dots, f_n, g_1, \dots, g_k)$. Thus it suffices to show that this spanning list is linearly independent.

Suppose it is not. Then we can write

$$0 = a_1 e_1 + \cdots + a_m e_m + b_1 f_1 + \cdots + b_n f_n + c_1 g_1 + \cdots + c_k g_k,$$

where at least one of the $b$'s and $c$'s is not zero. Say without loss of generality that,

$$a_1 e_1, \dots, a_m e_m, b_1 f_1, \dots, b_n f_n \in \operatorname{span}(g_1, \dots, g_k).$$

where at least one of the $b$'s is not zero. Thus

$$\begin{aligned} a_1 e_1, \dots, a_m e_m, b_1 f_1, \dots, b_n f_n \in V_1 \cap V_2 &= \operatorname{span}(e_1, \dots, e_m) \\ &= c_1 e_1, \dots, c_m e_m, \end{aligned}$$

contradicting that $e_1, \dots, e_m, f_1, \dots, f_n$ is a basis of $V_1$ (cf. 2.15.) ⊣

## Exercises

**Exercise 1.** Show that the subspaces of $\mathbb{R}^2$ are precisely $\{0\}$, all lines in $\mathbb{R}^2$ containing the origin, and $\mathbb{R}^2$.

We can prove that $\mathbb{R}^2$ has dimension 2. By 2.22, the subspaces of $\mathbb{R}^2$ can have dimensions 0 (indicating an empty spanning list, thus $\{0\}$), 1 (lines), and 2 ($\mathbb{R}^2$ itself by 2.24).

Chapter 3

# Linear Maps

## 3A Vector Space of Linear Maps

**Definition 3.1.** A *linear map* from $V$ to $W$ is a homomorphism of $V$ into $W$.

**Remark 3.2.** The class of vector spaces over a fixed field is an elementasy class with regard to the sentences expressing 1.3 within a suitable first-order language $\mathfrak{L}$, say, a two-sorted (scalars and vectors) one with addition and scalar multiplication.
A homomorphism in 3.1 is a function of one sturcture into another, preserving predicates (none here) and functions (addition and scalar multiplications) in $\mathfrak{L}$.
Note also that just like scalar multiplication functions, linear maps are often denoted by $Tv$, which just means $T(v)$.
The set of linear maps from $V$ to $W$ is denoted by $\mathcal{L}(V, W)$, and $\mathcal{L}(V, V)$ can be written as $\mathcal{L}(V)$.

**Example 3.3.** Here are some examples of linear maps.

1. Differentiation is a linear map. For example consider polynomials (2.6), define $D \in \mathcal{L}(\mathcal{P}(\mathbb{R}))$ by $Dp = p'$. This indeed formally states the linearity of differentiation. So is integration.
2. Let $m$ and $n$ be positive integers, let $A_{j,k} \in \mathbb{F}$ for each $j = i, \dots, m$ and each $k = 1, \dots, n$, and a linear map $T \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$ can be defined by

$$T(x_1, \dots, x_n) = (A_{1,1}x_1 + \cdots + A_{1,n}x_n, \dots, A_{m,1}x_1 + \cdots + A_{m,n}x_n).$$

   Actually every linear map from $\mathbb{F}^n$ to $\mathbb{F}^m$ is of this form.
3. Fix a polynomial $q \in \mathcal{P}(\mathbb{R})$. A linear map $T \in \mathcal{L}(\mathcal{P}(\mathbb{R}))$ can be defined by

$$(Tp)(x) = p(q(x)).$$

**Theorem 3.4.** *Suppose $v_1, \dots, v_n$ is a basis of $V$ and $w_1, \dots, w_n \in W$. Then there exists a unique linear map $T : V \to W$ such that*

$$Tv_k = w_k$$

*for each $k = 1, \dots, n$.*

*Proof.* Given $w_1, \dots, w_n \in W$, a linear map is natually defined. For each $v \in V$, where $v = \sum_{i=1}^n a_i v_i$ and $a_i \in \mathbb{F}$ for $i = 1, \dots, n$, we have that $Tv = \sum_{i=1}^n a_i w_i$, which is obviously unique. $\dashv$

*Comment.* The uniqueness means that a linear map is completely determined by its values on a basis.

**Definition 3.5.** Suppose $S, T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbb{F}$. The *sum* $S + T$ and the *product* $\lambda T$ are the linear maps from $V$ to $W$ defined by

$$(S + T)(v) = Sv + Tv \quad \text{and} \quad (\lambda T)(v) = \lambda(Tv)$$

for all $v \in V$. $\mathcal{L}(V, W)$ is then a vector space.

**Definition 3.6.** If $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, then the *product* $ST \in \mathcal{L}(U, W)$ is defined by

$$(ST)(u) = S(Tu)$$

for all $u \in U$.

Note that the product of linear maps satisfies associativity, identity and distributive properties. Not commutative though.

**Definition 3.7.** Suppose $T$ is a linear map from $V$ to $W$. Then $T(0) = 0$.

*Proof.* The linear map defined in 3.4 makes this immediate. Or, consider that

$$V(0) = V(0 + 0) = V(0) + V(0).$$

Add the additive inverse of $T(0)$ to each side of the equation above to conclude that $T(0) = 0$. $\dashv$

## 3B Null Spaces and Ranges

**Definition 3.8.** For $T \in \mathcal{L}(V, W)$, the *null space* (or *kernel*) of $T$, denoted by null $T$, is the subset of $V$ consisting of those vectors that $T$ maps to 0:

**Theorem 3.9.** *Suppose $T \in \mathcal{L}(V, W)$. Then* null $T$ *is a subspace of $V$.*

**Definition 3.10.** A function $T : V \to W$ is called *injective* or *one-to-one* if $Tu = Tv$ implies $u = v$.

**Theorem 3.11.** *Let $T \in \mathcal{L}(V, W)$. Then $T$ is injective iff* null $T = \{0\}$.

**Definition 3.12.** For $T \in \mathcal{L}(V, W)$, the *range* of $T$ is the subset of $W$ consisting of those vectors that are equal to $Tv$ for some $v \in V$.

**Theorem 3.13.** *If $T \in \mathcal{L}(V, W)$, then* range $T$ *is a subspace of $W$.*

**Definition 3.14.** A function $T : V \to W$ is called *surjective* or *onto* $W$ if its range equals $W$.

**Theorem 3.15.** ***Fundamental Theorem of Linear Maps*** *Suppose $V$ is finite-dimensional and $T \in \mathcal{L}(V, W)$. Then* range $T$ *is finite-dimensional and*

$$\dim V = \dim \operatorname{null} T + \dim \operatorname{range} T.$$

*Proof.* Say that the dimension of $V$ is $n$ and the dimension of null $T$ is $m$. By 3.9 $m \le n$. Let $v_1, \dots, v_m$ be a basis of null $T$ and expend it by 2.18 to a basis of $V$ $v_1, \dots, v_n$. Note that range $T$ is spanned by $Tv_1, \dots, Tv_n$, thus also by $Tv_{m+1}, \dots, Tv_n$. Then $\dim \operatorname{range} T \le n - m$. Suppose that $\dim \operatorname{range} T < n - m$ when $n > m$. Then $Tv_{m+1}, \dots, Tv_n$ is linearly dependent, thus we have

$$Tv_n = a_1 Tv_{m+1}, \dots, a_{n-m-1} Tv_{n-1},$$

where $a_i \in \mathbb{F}$ for $i = 1, \dots, n-m-1$, when $n - m > 1$. That is,

$$v_n - \sum_{i=1}^{n-m-1} a_i v_{m+i} \in \operatorname{null} T.$$

when $n-m > 1$ and $v_n \in \operatorname{null} T$ when $n-m = 1$, from which we can derive that $v_1, \dots, v_n$ is linearly dependent, hence contradiction. Therefore, $\dim \operatorname{range} T = n - m$, as desired. ⊣

**Theorem 3.16.** *Suppose $V$ and $W$ are finite-dimensional vector spaces such that $\dim V > \dim W$. Then no linear map from $V$ to $W$ is injective.*

*Proof.* Trivial by 2.22, 3.11, and 3.13. ⊣

**Theorem 3.17.** *Suppose $V$ and $W$ are finite-dimensional vector spaces such that $\dim V < \dim W$. Then no linear map from $V$ to $W$ is surjective.*

**Remark 3.18.** Fix positive integers $m$ and $n$, and let $A_{j,k} \in \mathbb{F}$ for $j = 1, \dots, m$ and $k = 1, \dots, n$. Consider the homogeneous (the constant term on the right side of each equation below is 0) system of linear equations

$$\begin{gathered} \sum_{k=1}^{n} A_{1,k} x_k = 0 \\ \vdots \\ \sum_{k=1}^{n} A_{m,k} x_k = 0. \end{gathered}$$

Clearly $x_1 = \cdots = x_n = 0$ is a solution of the system of equations above; the question here is whether any other solutions exist.
Define $T : \mathbb{F}^n \to \mathbb{F}^m$ by

$$T(x_1, \dots, x_n) = \left( \sum_{k=1}^{n} A_{1,k} x_k, \ \dots, \ \sum_{k=1}^{n} A_{m,k} x_k \right).$$

The equation $T(x_1, \dots, x_n) = 0$ is the same as the homogeneous system of linear equations above. Thus we want to know if $\operatorname{null} T$ is strictly bigger than $\{0\}$.

**Theorem 3.19.** *A homogeneous system of linear equations with more variables than equations has nonzero solutions.*

**Theorem 3.20.** *An inhomogeneous system of linear equations with more equations than variables has no solution for some choice of the constant terms.*

## 3C Matrices

**Remark 3.21.** Suppose $A$ is a matrix. The notation $A_{j,k}$ denotes the entry in row $j$, column $k$ of $A$.

**Definition 3.22.** Suppose $T \in \mathcal{L}(V, W)$ and $v_1, \dots, v_n$ is a basis of $V$ and $w_1, \dots, w_m$ is a basis of $W$. The *matrix* of $T$ with respect to these bases is the $m$-by-$n$ matrix $\mathcal{M}(T)$ whose entries $A_{j,k}$ are defined by

$$Tv_k = A_{1,k}w_1 + \cdots + A_{m,k}w_m.$$

If the bases $v_1, \dots, v_n$ and $w_1, \dots, w_m$ are not clear from the context, then the notation

$$\mathcal{M}\big(T, (v_1, \dots, v_n), (w_1, \dots, w_m)\big)$$

is used.

If you think of elements of $\mathbb{F}^m$ as columns of $m$ numbers, then you can think of the $k^{\text{th}}$ column of as $\mathcal{M}(T)$ applied to the $k^{\text{th}}$ standard basis vector.

**Example 3.23.** Suppose $D \in \mathcal{L}(\mathcal{P}_3(\mathbb{R}), \mathcal{P}_2(\mathbb{R}))$ is the differentiation map defined by $Dp = p'$. Because $(x^n)' = nx^{n-1}$, the matrix of $D$ with respect to the standard bases is the 3-by-4 matrix below:

$$\mathcal{M}(D) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

**Theorem 3.24.** *Suppose $\lambda \in \mathbb{F}$ and $S, T \in \mathcal{L}(V, W)$. Then $\mathcal{M}(S + T) = \mathcal{M}(S) + \mathcal{M}(T)$ and $\mathcal{M}(\lambda T) = \lambda\mathcal{M}(T)$.*

**Theorem 3.25.** *For $m$ and $n$ positive integers, the set of all $m$-by-$n$ matrices with entries in $\mathbb{F}$, denoted by $\mathbb{F}^{m,n}$, is a vector space of dimension $mn$.*

Matrix multiplication is defined with respect to the composition of linear maps.

**Theorem 3.26.** *If $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, then $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$.*

Below are two other ways to think of matrix multiplication.

**Theorem 3.27.** *Suppose $A$ is an $m$-by-$n$ matrix and $B$ is an $n$-by-$p$ matrix. Then*

$$(AB)_{j,k} = A_{j,\cdot} B_{\cdot,k}$$

*and*

$$(AB)_{\cdot,k} = AB_{\cdot,k}$$

*if $1 \leq j \leq m$ and $1 \leq k \leq p$.*

**Theorem 3.28.** *Suppose $A$ is an $m$-by-$n$ matrix and $b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$ is an $n$-by-$1$ matrix. Then*

$$Ab = b_1 A_{\cdot,1} + \cdots + b_n A_{\cdot,n}.$$

**Theorem 3.29.** *Suppose $C$ is an $m$-by-$c$ matrix and $R$ is a $c$-by-$n$ matrix.*

*(a) If $k \in \{1, \dots, n\}$, then column $k$ of $CR$ is a linear combination of the columns of $C$, with the coefficients of this linear combination coming from column $k$ of $R$.*

*(b) If $j \in \{1, \dots, m\}$, then row $j$ of $CR$ is a linear combination of the rows of $R$, with the coefficients of this linear combination coming from row $j$ of $C$.*

*Proof.* (a) is by 3.27 and 3.28. (b) can similarly proved. ⊣

**Definition 3.30.** Suppose $A$ is an $m$-by-$n$ matrix with entries in $\mathbb{F}$.

(a) The *column rank* of $A$ is the dimension of the span of the columns of $A$ in $\mathbb{F}^{m,1}$.

(b) The *row rank* of $A$ is the dimension of the span of the rows of $A$ in $\mathbb{F}^{1,n}$.

Suppose $A$ is an $m$-by-$n$ matrix. Note that the transpose $A \mapsto A^\top \in \mathcal{L}(\mathbb{F}^{m,n}, \mathbb{F}^{n,m})$. Note also that the column and row ranks of $A$ are at most $\min\{m, n\}$.

**Theorem 3.31.** *Suppose $A$ is an $m$-by-$n$ matrix with entries in $\mathbb{F}$ and column rank $c \geq 1$. Then there exist an $m$-by-$c$ matrix $C$ and a $c$-by-$n$ matrix $R$, both with entries in $\mathbb{F}$, such that $A = CR$.*

*Proof.* Reduce the columns of $A$ to a basis of length $c$, making $C \in \mathbb{F}^{m,c}$. Each column of $A$ is a linear combination of the columns of $C$, which gives $R$ by 3.29 (a). ⊣

**Theorem 3.32.** *The column rank of $A \in \mathbb{F}^{m,n}$ equals its row rank.*

*Proof.* By 3.29 (b) and 3.31 the row rank of $A$ is at most $c$. Similarly the column rank is at most row rank, giving the result. ⊣

Cf. 3.133 to do and 7A.8

Both are called *rank*.

## 3D Invertibility and Isomorphisms

**Definition 3.33.** A linear map $T \in \mathcal{L}(V, W)$ is called *invertible* if there exists a linear map $S \in \mathcal{L}(W, V)$ (called *inverse*) such that $ST = I$ and $TS = I$.

Note that such an inverse is unique, and is denoted by $T^{-1}$.

**Theorem 3.34.** *A linear map is invertible if and only if it is injective and surjective.*

**Theorem 3.35.** *Suppose that $V$ and $W$ are finite-dimensional vector spaces,* $\dim V = \dim W$, *and $T \in \mathcal{L}(V, W)$. Then*

$$T \text{ is invertible} \iff T \text{ is injective} \iff T \text{ is surjective.}$$

This is almost immediate by 3.15.

**Theorem 3.36.** *Suppose $V$ and $W$ are finite-dimensional vector spaces of the same dimension, $S \in \mathcal{L}(V, W)$, and $T \in \mathcal{L}(W, V)$. Then*

$$ST = I \quad \text{if and only if} \quad TS = I.$$

*Proof.* Suppose $ST = I$. If $v \in V$ and $Tv = 0$, then

$$v = (ST)v = S(Tv) = 0.$$

Thus $T$ is injective by 3.11 and invertible by 3.35.

Now multiply both sides of the equation $ST = I$ with $T^{-1}$, giving

$$S = T^{-1}.$$

Thus $TS = TT^{-1} = I$, as desired. The other direction is similar. ⊣

**Remark 3.37.** For sure, an invertible linear map is an *isomorphism*. Cf. 3.1.

**Theorem 3.38.** *Two finite-dimensional vector spaces over $\mathbb{F}$ are isomorphic iff they have the same dimension.*

*Proof.* (⇒) Immediate by 3.15. (⇐) Define a map that maps vectors in one basis to those in the other. ⊣

**Theorem 3.39.** *Suppose $v_1, \dots, v_n$ is a basis of $V$ and $w_1, \dots, w_m$ is a basis of $W$. Then $\mathcal{M}$ is an isomorphism between $\mathcal{L}(V, W)$ and $\mathbb{F}^{m,n}$.*

*Proof.* $\mathcal{M}$ is linear by 3.24. If $T \in \mathcal{L}(V, W)$ and $\mathcal{M}(T) = 0$, then $Tv_k = 0$ for each $k = 1, \dots, n$ by 3.21. Thus $T = 0$. Thus $\mathcal{M}$ is injective by 3.11. The surjectivity is by 3.4. ⊣

**Corollary 3.40.** *Suppose $V$ and $W$ are finite-dimensional. Then*

$$\dim \mathcal{L}(V, W) = (\dim V)(\dim W).$$

Vectors has matrices.

**Theorem 3.41.** *Suppose $T \in \mathcal{L}(V, W)$ and $v_1, \dots, v_n$ is a basis of $V$ and $w_1, \dots, w_m$ is a basis of $W$. Let $1 \leq k \leq n$. Then the $k^{th}$ column of $\mathcal{M}(T)$, which is denoted by $\mathcal{M}(T)_{\cdot,k}$, equals $\mathcal{M}(Tv_k)$.*

Linear maps act like matrix multiplication.

**Theorem 3.42.** *Suppose $T \in \mathcal{L}(V, W)$ and $v \in V$. Suppose $v_1, \dots, v_n$ is a basis of $V$ and $w_1, \dots, w_m$ is a basis of $W$. Then*

$$\mathcal{M}(Tv) = \mathcal{M}(T)\mathcal{M}(v).$$

**Theorem 3.43.** *Suppose $V$ and $W$ are finite-dimensional and $T \in \mathcal{L}(V, W)$. Then* $\dim \operatorname{range} T$ *equals the column rank of $\mathcal{M}(T)$.*

**Definition 3.44.** A square matrix $A$ is called *invertible* or *nonsingular* if there is a square matrix $B$ of the same size such that $AB = BA = I$; we call $B$ the *inverse* of $A$ and denote it by $A^{-1}$.

**Theorem 3.45.** *Suppose that $u_1, \dots, u_n$ and $v_1, \dots, v_n$ are bases of $V$. Then the matrices*

$$\mathcal{M}\big(I, (u_1, \dots, u_n), (v_1, \dots, v_n)\big) \quad \text{and} \quad \mathcal{M}\big(I, (v_1, \dots, v_n), (u_1, \dots, u_n)\big)$$

*are invertible, and each is the inverse of the other.*

**Theorem 3.46.** *Suppose $T \in \mathcal{L}(V)$. Suppose $u_1, \dots, u_n$ and $v_1, \dots, v_n$ are bases of $V$. Let*

$$A = \mathcal{M}\big(T, (u_1, \dots, u_n)\big) \quad \text{and} \quad B = \mathcal{M}\big(T, (v_1, \dots, v_n)\big)$$

*and*

$$C = \mathcal{M}\big(I, (u_1, \dots, u_n), (v_1, \dots, v_n)\big).$$

*Then*

$$A = C^{-1}BC.$$

**Theorem 3.47.** *Suppose that $v_1, \dots, v_n$ is a basis of $V$ and $T \in \mathcal{L}(V)$ is invertible. Then*

$$\mathcal{M}(T^{-1}) = \big(\mathcal{M}(T)\big)^{-1},$$

*where both matrices are with respect to the basis $v_1, \dots, v_n$.*

## Exercises

**Exercise 11.** Suppose $V$ is finite-dimensional and $S, T \in \mathcal{L}(V)$. Prove that

$$ST \text{ is invertible} \iff S \text{ and } T \text{ are invertible.}$$

# 3F Duality

**Definition 3.48.** A *linear functional* on $V$ is a linear map from $V$ to $\mathbb{F}$.

**Definition 3.49.** The *dual space* of $V$, denoted by $V'$, is the vector space of all linear functionals on $V$.

Chapter 4

# Polynomials

**Definition 4.1.** Suppose $z \in \mathbb{C}$.
The *complex conjugate* of $z \in \mathbb{C}$, denoted by $\bar{z}$, is defined by

$$\bar{z} = \operatorname{Re} z - (\operatorname{Im} z)i.$$

The *absolute value* of a complex number $z$, denoted by $|z|$, is defined by

$$|z| = \sqrt{(\operatorname{Re} z)^2 + (\operatorname{Im} z)^2}.$$

**Remark 4.2.** Suppose $w, z \in \mathbb{C}$. Then

(a) $z\bar{z} = |z|^2$.
(b) $\overline{wz} = \bar{w}\bar{z}$.
(c) $|wz| = |w||z|$.
(d) $|w + z| \leq |w| + |z|$.

**Theorem 4.3.** *Suppose $m$ is a positive integer and $p \in \mathcal{P}(\mathbb{F})$ is a polynomial of degree $m$. Suppose $\lambda \in \mathbb{F}$. Then $p(\lambda) = 0$ if and only if there exists a polynomial $q \in \mathcal{P}(\mathbb{F})$ of degree $m - 1$ such that*

$$p(z) = (z - \lambda)q(z)$$

*for every $z \in \mathbb{F}$.*

*Proof.* Suppose $p(\lambda) = 0$. Let $a_0, a_1, \ldots, a_m \in \mathbb{F}$ be such that

$$p(z) = a_0 + a_1 z + \cdots + a_m z^m$$

for all $z \in \mathbb{F}$. Then

$$p(z) = p(z) - p(\lambda) = a_1(z - \lambda) + \cdots + a_m(z^m - \lambda^m)$$

for all $z \in \mathbb{F}$. For each $k \in \{1, \dots, m\}$, we have that

$$z^k - \lambda^k = (z - \lambda) \sum_{j=1}^{k} \lambda^{j-1} z^{k-j}$$

as desired. The other direction is trivial. ⊣

**Theorem 4.4.** *Suppose $m$ is a positive integer and $p \in \mathcal{P}(\mathbb{F})$ is a polynomial of degree $m$. Then $p$ has at most $m$ zeros in $\mathbb{F}$.*

This is by induction and 4.3.

**Corollary 4.5.** *The coefficients of a polynomial are uniquely determined by the polynomial.*

*Proof.* If a polynomial had two different sets of coefficients, then subtracting the two representations of the polynomial would give a polynomial with some nonzero coefficients but infinitely many zeros. ⊣

In particular, the degree of a polynomial is uniquely defined.

**Theorem 4.6.** *Suppose that $p, s \in \mathcal{P}(\mathbb{F})$, with $s \neq 0$. Then there exist unique polynomials $q, r \in \mathcal{P}(\mathbb{F})$ such that*

$$p = sq + r$$

*and* $\deg r < \deg s$.

*Proof.* Let $n = \deg p$ and let $m = \deg s$. If $n < m$, then take $q = 0$ and $r = p$ to get the desired equation $p = sq + r$ with $\deg r < \deg s$. Thus we now assume that $n \geq m$. The list

$$1, z, \dots, z^{m-1}, s, zs, \dots, z^{n-m}s$$

is linearly independent in $\mathcal{P}_n(\mathbb{F})$ because each polynomial in this list has a different degree. Also, this list has length $n + 1$, which equals $\dim \mathcal{P}_n(\mathbb{F})$. Hence it is a basis of $\mathcal{P}_n(\mathbb{F})$. Because $p \in \mathcal{P}_n(\mathbb{F})$ and the above list is a basis of $\mathcal{P}_n(\mathbb{F})$, there exist unique constants $a_0, a_1, \dots, a_{m-1} \in \mathbb{F}$ and $b_0, b_1, \dots, b_{n-m} \in \mathbb{F}$ such that

$$\begin{aligned} p &= a_0 + a_1 z + \cdots + a_{m-1} z^{m-1} + b_0 s + b_1 zs + \cdots + b_{n-m} z^{n-m} s \\ &= \underbrace{a_0 + a_1 z + \cdots + a_{m-1} z^{m-1}}_{r} + s \underbrace{\left(b_0 + b_1 z + \cdots + b_{n-m} z^{n-m}\right)}_{q}. \end{aligned}$$

With $r$ and $q$ as defined above, we see that $p$ can be written as $p = sq + r$ with $\deg r < \deg s$, as desired. The uniqueness of $q, r \in \mathcal{P}(\mathbb{F})$ satisfying these conditions follows from the uniqueness of the constants $a_0, a_1, \ldots, a_{m-1} \in \mathbb{F}$ and $b_0, b_1, \ldots, b_{n-m} \in \mathbb{F}$. ⊣

**Theorem 4.7.** ***Fundamental Theorem of Algebra,*** $1^{st}$ ***Version***
*Every nonconstant polynomial with complex coefficients has a zero in* $\mathbb{C}$.

*Proof.* to do ⊣

**Theorem 4.8.** ***Fundamental Theorem of Algebra,*** $2^{nd}$ ***Version***
*If* $p \in \mathcal{P}(\mathbb{C})$ *is a nonconstant polynomial, then* $p$ *has a unique factorization (except for the order of the factors) of the form*

$$p(z) = c(z - \lambda_1) \cdots (z - \lambda_m),$$

*where* $c, \lambda_1, \ldots, \lambda_m \in \mathbb{C}$.

*Proof.* The existence of the desired factorization is by induction, using 4.7 and 4.3. The uniqueness of the desired factorization is again by induction: If

$$(z - \lambda_1) \cdots (z - \lambda_m) = (z - \tau_1) \cdots (z - \tau_m)$$

for all $z \in \mathbb{C}$, then because the left side of the equation above equals 0 when $z = \lambda_1$, one of the $\tau$'s on the right side equals $\lambda_1$. Relabeling, we can assume that $\tau_1 = \lambda_1$. If $z \neq \lambda_1$, divide both sides of the eqation by $z - \lambda_1$, getting

$$(z - \lambda_2) \cdots (z - \lambda_m) = (z - \tau_2) \cdots (z - \tau_m)$$

for all $z \in \mathbb{C}$ except possibly $z = \lambda_1$. Actually for all $z \in \mathbb{C}$, because otherwise by subtracting the right side from the left side we would get a nonzero polynomial that has infinitely many zeros. By inductive hypothesis we are done. ⊣

**Theorem 4.9.** *Suppose* $p \in \mathcal{P}(\mathbb{C})$ *is a polynomial with real coefficients. If* $\lambda \in \mathbb{C}$ *is a zero of* $p$, *then so is* $\bar{\lambda}$.

*Proof.* Take the complex conjugate of both sides of the equation. By 4.2 (b) we are done. ⊣

**Theorem 4.10.** *Suppose* $p \in \mathcal{P}(\mathbb{R})$ *is a nonconstant polynomial. Then* $p$ *has a unique factor-*

*ization (except for the order of the factors) of the form*

$$p(x) = c(x - \lambda_1) \cdots (x - \lambda_m)(x^2 + b_1 x + c_1) \cdots (x^2 + b_M x + c_M),$$

*where* $c, \lambda_1, \ldots, \lambda_m, b_1, \ldots, b_M, c_1, \ldots, c_M \in \mathbb{R}$, *with* $b_k^2 < 4c_k$ *for each* $k$.

*Proof.* Existence: Think of $p$ as an element of $\mathcal{P}(\mathbb{C})$. If all (complex) zeros of $p$ are real, then we have the desired factorization by 4.8. Thus suppose $p$ has a zero $\lambda \in \mathbb{C}$ with $\lambda \notin \mathbb{R}$. By 4.9, $\bar{\lambda}$ is a zero of $p$. Thus we can write

$$p(x) = (x - \lambda)(x - \bar{\lambda})q(x) = \left(x^2 - 2(\operatorname{Re}\lambda)x + |\lambda|^2\right)q(x)$$

for some polynomial $q \in \mathcal{P}(\mathbb{C})$ of degree two less than the degree of $p$. If we can prove that $q$ has real coefficients, then using induction on the degree of $p$ completes the proof of the existence part of this result.

To prove that $q$ has real coefficients, we solve the equation above for $q$, getting

$$q(x) = \frac{p(x)}{x^2 - 2(\operatorname{Re}\lambda)x + |\lambda|^2}$$

for all $x \in \mathbb{R}$. The equation above implies that $q(x) \in \mathbb{R}$ for all $x \in \mathbb{R}$. Writing

$$q(x) = a_0 + a_1 x + \cdots + a_{n-2} x^{n-2},$$

where $n = \deg p$ and $a_0, \ldots, a_{n-2} \in \mathbb{C}$, we thus have

$$0 = \operatorname{Im} q(x) = (\operatorname{Im} a_0) + (\operatorname{Im} a_1)x + \cdots + (\operatorname{Im} a_{n-2})x^{n-2}$$

for all $x \in \mathbb{R}$. This implies that $\operatorname{Im} a_0, \ldots, \operatorname{Im} a_{n-2}$ all equal 0 (by 4.4). Thus all coefficients of $q$ are real, as desired. Hence the desired factorization exists.

Uniqueness: A factor of $p$ of the form $x^2 + b_k x + c_k$ with $b_k^2 < 4c_k$ can be uniquely written as $(x - \lambda_k)(x - \bar{\lambda}_k)$ with $\lambda_k \in \mathbb{C}$. Note that two different factorizations of $p$ as an element of $\mathcal{P}(\mathbb{R})$ would lead to two different factorizations of $p$ as an element of $\mathcal{P}(\mathbb{C})$, contradicting 4.8. ⊣

Chapter 5

# Eigenvalues and Eigenvectors

## 5A Invariant Subspaces

**Definition 5.1.** A linear map from a vector space to itself is called an *operator*.

Suppose $T \in \mathcal{L}(V)$. If $m \geq 2$ and $V = V_1 \oplus \cdots \oplus V_m$, where each $V_k$ is a nonzero subspace of $V$. We natually want to study $T|_{V_k}$. If would be useful if $T|_{V_k}$ is an operator on $V_k$. Hence the following definition.

**Definition 5.2.** Suppose $T \in \mathcal{L}(V)$. A subspace $U$ of $V$ is called *invariant* under $T$ if $Tu \in U$ for every $u \in U$.

Note that $\{0\}$, $V$, null $T$, and range $T$ are invariant subspaces of $V$.

Take any $v \in V$ and $v \neq 0$. We want to verify if span$(v)$ is an invariant subspace.

**Definition 5.3.** Suppose $T \in \mathcal{L}(V)$. A number $\lambda \in \mathbb{F}$ is called an *eigenvalue* of $T$ if there exists $v \in V$ such that $v \neq 0$ and $Tv = \lambda v$.

**Theorem 5.4.** *Cf. 5.3. The following are equivalent if $V$ is finite-dimensional.*

*(a) $\lambda$ is an eigenvalue of $T$.*
*(b) $T - \lambda I$ is not injective.*
*(c) $T - \lambda I$ is not surjective.*
*(d) $T - \lambda I$ is not invertible.*

**Definition 5.5.** Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$ is an eigenvalue of $T$. A vector $v \in V$ is called an *eigenvector* of $T$ corresponding to $\lambda$ if $v \neq 0$ and $Tv = \lambda v$.

**Remark 5.6.** $Tv = \lambda v$ iff $(T - \lambda I)v = 0$, thus $v \in V$ with $v \neq 0$ is an eigenvector of $T$ corresponding to $\lambda$ iff $v \in \operatorname{null}(T - \lambda I)$.

**Theorem 5.7.** *Suppose $T \in \mathcal{L}(V)$. Then every list of eigenvectors of $T$ corresponding to distinct eigenvalues of $T$ is linearly independent.*

*Proof.* Suppose the desired result is false. Then there exists a smallest positive integer $m$ such that there exists a linearly dependent list $v_1, \dots, v_m$ of eigenvectors of $T$ corresponding to distinct eigenvalues $\lambda_1, \dots, \lambda_m$ of $T$ (note that $m \geq 2$ because an eigenvector is, by definition, nonzero). Thus there exist $a_1, \dots, a_m \in \mathbb{F}$, none of which are 0 (because of the minimality of $m$), such that

$$a_1 v_1 + \cdots + a_m v_m = 0.$$

Apply $T - \lambda_m I$ to both sides of the equation above, getting

$$a_1(\lambda_1 - \lambda_m)v_1 + \cdots + a_{m-1}(\lambda_{m-1} - \lambda_m)v_{m-1} = 0.$$

Because the eigenvalues $\lambda_1, \dots, \lambda_m$ are distinct, none of the coefficients above equal 0. Thus $v_1, \dots, v_{m-1}$ is a linearly dependent list of $m - 1$ eigenvectors of $T$ corresponding to distinct eigenvalues, contradicting the minimality of $m$. This contradiction completes the proof. $\dashv$

*Comment.* Can be proved by induction. But this treatment is way more elegant.

**Corollary 5.8.** *Suppose $V$ is finite-dimensional. Then each operator on $V$ has at most $\dim V$ distinct eigenvalues.*

**Definition 5.9.** Suppose $T \in \mathcal{L}(V)$ and $m$ is a positive integer.

- $T^m \in \mathcal{L}(V)$ is defined by
$$T^m = \underbrace{T \cdots T}_{m \text{ times}}.$$
- $T^0$ is defined to be the identity operator $I$ on $V$.
- If $T$ is invertible with inverse $T^{-1}$, then $T^{-m} \in \mathcal{L}(V)$ is defined by
$$T^{-m} = (T^{-1})^m.$$

**Definition 5.10.** Suppose $T \in \mathcal{L}(V)$ and $p \in \mathcal{P}(\mathbb{F})$ is a polynomial given by

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_m z^m$$

for all $z \in \mathbb{F}$. Then $p(T)$ is the operator on $V$ defined by

$$p(T) = a_0 I + a_1 T + a_2 T^2 + \cdots + a_m T^m.$$

Note that $p \mapsto p(T)$ is linear with a fixed operator.

**Remark 5.11.** Suppose $p, q \in \mathcal{P}(\mathbb{F})$ and $T \in \mathcal{L}(V)$. Then

(a) $(pq)(T) = p(T)q(T)$;
(b) $p(T)q(T) = q(T)p(T)$.

This is by computation.

**Theorem 5.12.** *Suppose $T \in \mathcal{L}(V)$ and $p \in \mathcal{P}(\mathbb{F})$. Then* null $p(T)$ *and* range $p(T)$ *are invariant under $T$.*

*Proof.* Suppose $u \in \operatorname{null} p(T)$. Then $p(T)u = 0$. Thus by 5.11 (b)

$$(p(T))(Tu) = T(p(T)u) = T(0) = 0.$$

Hence $Tu \in \operatorname{null} p(T)$. Thus null $p(T)$ is invariant under $T$, as desired.

Suppose $u \in \operatorname{range} p(T)$. Then there exists $v \in V$ such that $u = p(T)v$. Thus by 5.11 (b)

$$Tu = T(p(T)v) = p(T)(Tv).$$

Hence $Tu \in \operatorname{range} p(T)$. Thus range $p(T)$ is invariant under $T$, as desired. ⊣

## 5B The Minimal Polynomial

**Theorem 5.13.** *Every operator on a finite-dimensional nonzero complex vector space has an eigenvalue.*

*Proof.* Suppose $V$ is a finite-dimensional complex vector space of dimension $n > 0$ and $T \in \mathcal{L}(V)$. Choose $v \in V$ with $v \neq 0$. Then

$$v,\ Tv,\ T^2 v,\ \ldots,\ T^n v$$

is not linearly independent, because $V$ has dimension $n$ and this list has length $n+1$. Hence some linear combination (with not all the coefficients equal to 0) of the vectors above equals 0. Thus there exists a nonconstant polynomial $p$ of smallest degree such that

$$p(T)v = 0.$$

By 4.7 there exists $\lambda \in \mathbb{C}$ such that $p(\lambda) = 0$. Hence there exists a polynomial $q \in \mathcal{P}(\mathbb{C})$ such that

$$p(z) = (z - \lambda)q(z)$$

for every $z \in \mathbb{C}$. This implies that

$$0 = p(T)v = (T - \lambda I)(q(T)v).$$

Because $q$ has smaller degree than $p$, we know that $q(T)v \neq 0$. Thus the equation above implies that $\lambda$ is an eigenvalue of $T$ with eigenvector $q(T)v$. ⊣

**Definition 5.14.** A *monic polynomial* is a polynomial whose highest-degree coefficient equals 1.

**Theorem 5.15.** *Suppose $V$ is finite-dimensional and $T \in \mathcal{L}(V)$. Then there is a unique monic polynomial $p \in \mathcal{P}(\mathbb{F})$ of smallest degree such that $p(T) = 0$. Furthermore,* $\deg p \leq \dim V$.

*Proof.* If $\dim V = 0$, then $I$ is the zero operator on $V$ and thus we take $p$ to be the constant polynomial 1.

Now use induction on $\dim V$. Assume that $\dim V > 0$ and that the desired result is true for all operators on all vector spaces over $\mathbb{F}$ of smaller dimensions. Let $v \in V$ be such that $v \neq 0$. The list $v, Tv, \dots, T^{\dim V}v$ has length $1 + \dim V$ and thus is linearly dependent. By 2.11 there is a smallest positive integer $m \leq \dim V$ such that $T^m v$ is a linear combination of $v, Tv, \dots, T^{m-1}v$. Thus there exist scalars $c_0, c_1, c_2, \dots, c_{m-1} \in \mathbb{F}$ such that

$$c_0 v + c_1 Tv + \cdots + c_{m-1}T^{m-1}v + T^m v = 0. \tag{$*$}$$

Define a monic polynomial $q \in \mathcal{P}_m(\mathbb{F})$ by

$$q(z) = c_0 + c_1 z + \cdots + c_{m-1}z^{m-1} + z^m.$$

Then $(*)$ implies that $q(T)v = 0$. If $k$ is a nonnegative integer, then

$$q(T)(T^k v) = T^k(q(T)v) = T^k(0) = 0.$$

2.11 shows that $v, Tv, \ldots, T^{m-1}v$ is linearly independent. Thus the equation above implies that $\dim \operatorname{null} q(T) \geq m$. Hence

$$\dim \operatorname{range} q(T) = \dim V - \dim \operatorname{null} q(T) \leq \dim V - m.$$

Because range $q(T)$ is invariant under $T$, we can apply our induction hypothesis to the operator $T|_{\operatorname{range} q(T)}$ on the vector space range $q(T)$. Thus there is a monic polynomial $s \in \mathcal{P}(\mathbb{F})$ with

$$\deg s \leq \dim V - m \quad \text{and} \quad s(T|_{\operatorname{range} q(T)}) = 0.$$

Hence for all $v \in V$ we have

$$(sq)(T)(v) = s(T)(q(T)v) = 0$$

because $q(T)v \in \operatorname{range} q(T)$ and $s(T)|_{\operatorname{range} q(T)} = s(T|_{\operatorname{range} q(T)}) = 0$. Thus $sq$ is a monic polynomial such that $\deg sq \leq \dim V$ and $(sq)(T) = 0$.

Thus one of degree at most $\dim V$ exists. Hence so does one of the smallest degree. The uniqueness is by RAA, as usual. ⊣

**Definition 5.16.** Cf. 5.15. $p$ is called the *minimal polynomial* of $T$.

**Theorem 5.17.** *Suppose $V$ is finite-dimensional and $T \in \mathcal{L}(V)$.*

*(a) The zeros of the minimal polynomial of $T$ are the eigenvalues of $T$.*

*(b) If $V$ is a complex vector space, then the minimal polynomial of $T$ has the form*

$$(z - \lambda_1) \cdots (z - \lambda_m),$$

*where $\lambda_1, \ldots, \lambda_m$ is a list of all eigenvalues of $T$, possibly with repetitions.*

*Proof.* (a) First suppose $\lambda \in \mathbb{F}$ is a zero of $p$. Then $p$ can be written in the form

$$p(z) = (z - \lambda)q(z),$$

where $q$ is a monic polynomial with coefficients in $\mathbb{F}$. Because $p(T) = 0$, we have

$$0 = (T - \lambda I)(q(T)v)$$

for all $v \in V$. Because the degree of $q$ is less than the degree of the minimal polynomial $p$, there exists at least one vector $v \in V$ such that $q(T)v \neq 0$. The equation above thus implies that $\lambda$ is an eigenvalue of $T$, as desired.

For the other direction, suppose $\lambda \in \mathbb{F}$ is an eigenvalue of $T$. By $p(T)v = p(\lambda)v$ we are done.

(b) is by (a) and 4.8. ⊣

This with 5.15 and 4.4 should yield a second proof of 5.8.

**Remark 5.18.** Every monic polynomial is the minimal polynomial of some operator (described by the *companion matrix* of the polynomial, cf. 5B.16). Hence finding exact expressions for the zeros of a polynomial can be reduced to finding exact eigenvalues of an operator. The latter is thus not possible.

**Theorem 5.19.** *Suppose $V$ is finite-dimensional, $T \in \mathcal{L}(V)$, and $q \in \mathcal{P}(\mathbb{F})$. Then $q(T) = 0$ iff $q$ is a polynomial multiple of the minimal polynomial of $T$.*

*Proof.* Let $p$ denote the minimal polynomial of $T$. First suppose $q(T) = 0$. By 4.6 there exist polynomials $s, r \in \mathcal{P}(\mathbb{F})$ such that $q = ps + r$ and $\deg r < \deg p$. We have

$$0 = q(T) = p(T)s(T) + r(T) = r(T).$$

Note that $r = 0$ due to the minimality of $p$. Hence $q$ is a polynomial multiple of $p$, as desired. The other direction is trivial. ⊣

**Corollary 5.20.** *Suppose $V$ is finite-dimensional, $T \in \mathcal{L}(V)$, and $U$ is a subspace of $V$ that is invariant under $T$. Then the minimal polynomial of $T$ is a polynomial multiple of the minimal polynomial of $T|_U$.*

**Theorem 5.21.** *Suppose $V$ is finite-dimensional and $T \in \mathcal{L}(V)$. Then $T$ is not invertible if and only if the constant term of the minimal polynomial of $T$ is $0$.*

*Proof.* Suppose $T \in \mathcal{L}(V)$ and $p$ is the minimal polynomial of $T$. Then by 5.17 (a)

$$\begin{aligned} T \text{ is not invertible} &\Leftrightarrow 0 \text{ is an eigenvalue of } T \\ &\Leftrightarrow 0 \text{ is a zero of } p \\ &\Leftrightarrow \text{ the constant term of } p \text{ is } 0. \end{aligned}$$

⊣

**Lemma 5.22.** *Suppose $\mathbb{F} = \mathbb{R}$ and $V$ is finite-dimensional. Suppose also that $T \in \mathcal{L}(V)$ and $b, c \in \mathbb{R}$ with $b^2 < 4c$. Then $\dim \operatorname{null}(T^2 + bT + cI)$ is an even number.*

*Proof.* By replacing $V$ with $\operatorname{null}(T^2+bT+cI)$ and replacing $T$ with $T$ restricted to $\operatorname{null}(T^2+bT+cI)$, we can assume that $T^2 + bT + cI = 0$; we now need to prove that $\dim V$ is even. Suppose $\lambda \in \mathbb{R}$

and $v \in V$ are such that $Tv = \lambda v$. Then

$$0 = (T^2 + bT + cI)v = (\lambda^2 + b\lambda + c)v = \left(\left(\lambda + \frac{b}{2}\right)^2 + c - \frac{b^2}{4}\right)v.$$

The term in large parentheses above is a positive number. Thus the equation above implies that $v = 0$. Hence we have shown that $T$ has no eigenvectors.

Let $U$ be a subspace of $V$ that is invariant under $T$ and has the largest dimension among all subspaces of $V$ that are invariant under $T$ and have even dimension. If $U = V$, then we are done; otherwise assume there exists $w \in V$ such that $w \notin U$.

Let $W = \operatorname{span}(w, Tw)$. Then $W$ is invariant under $T$ because $T(Tw) = -bTw - cw$. Furthermore, $\dim W = 2$ because otherwise $w$ would be an eigenvector of $T$. Now

$$\dim(U + W) = \dim U + \dim W - \dim(U \cap W) = \dim U + 2,$$

where $U \cap W = \{0\}$ because otherwise $U \cap W$ would be a one-dimensional subspace of $V$ that is invariant under $T$ (impossible because $T$ has no eigenvectors).

Because $U + W$ is invariant under $T$, the equation above shows that there exists a subspace of $V$ invariant under $T$ of even dimension larger than $\dim U$. Thus the assumption that $U \neq V$ was incorrect. Hence $V$ has even dimension. $\dashv$

**Theorem 5.23.** *Operators on odd-dimensional vector spaces have eigenvalues.*

*Proof.* Suppose $\mathbb{F} = \mathbb{R}$ and $V$ is finite-dimensional. Let $n = \dim V$, and suppose $n$ is an odd number. Let $T \in \mathcal{L}(V)$. We will use induction on $n$ in steps of size two to show that $T$ has an eigenvalue. To get started, note that the desired result holds if $\dim V = 1$ because then every nonzero vector in $V$ is an eigenvector of $T$.

Now suppose that $n \geq 3$ and the desired result holds for all operators on all odd-dimensional vector spaces of dimension less than $n$. Let $p$ denote the minimal polynomial of $T$. If $p$ is a polynomial multiple of $x - \lambda$ for some $\lambda \in \mathbb{R}$, then $\lambda$ is an eigenvalue of $T$ by 5.17 (a) and we are done. Hence assume that there exist $b, c \in \mathbb{R}$ such that $b^2 < 4c$ and $p$ is a polynomial multiple of $x^2 + bx + c$ by 4.10.

There exists a monic polynomial $q \in \mathcal{P}(\mathbb{R})$ such that

$$p(x) = q(x)(x^2 + bx + c)$$

for all $x \in \mathbb{R}$. Now

$$0 = p(T) = (q(T))(T^2 + bT + cI),$$

which means that $q(T)$ equals 0 on $\operatorname{range}(T^2 + bT + cI)$. Because $\deg q < \deg p$ and $p$ is the minimal polynomial of $T$, this implies that $\operatorname{range}(T^2 + bT + cI) \neq V$.

By 3.15 we have that

$$\dim V = \dim \operatorname{null}(T^2 + bT + cI) + \dim \operatorname{range}(T^2 + bT + cI).$$

Because $\dim V$ is odd (by hypothesis) and $\dim \operatorname{null}(T^2 + bT + cI)$ is even by 5.22, the equation above shows that $\dim \operatorname{range}(T^2 + bT + cI)$ is odd.

Hence $\operatorname{range}(T^2 + bT + cI)$ is a subspace of $V$ that is invariant under $T$ and that has odd dimension less than $\dim V$. Our induction hypothesis now implies that $T$ restricted to $\operatorname{range}(T^2 + bT + cI)$ has an eigenvalue, which means that $T$ has an eigenvalue. ⊣

## Exercises

**Exercise 16.** Suppose $a_0, \dots, a_{n-1} \in \mathbb{F}$. Let $T$ be the operator on $\mathbb{F}^n$ whose matrix (with respect to the standard basis) is

$$\begin{pmatrix} 0 & & & & -a_0 \\ 1 & 0 & & & -a_1 \\ & 1 & \ddots & & -a_2 \\ & & \ddots & 0 & \vdots \\ & & & 1 & -a_{n-1} \end{pmatrix}.$$

Here all entries of the matrix are 0 except for all 1's on the line under the diagonal and the entries in the last column (some of which might also be 0). Show that the minimal polynomial of $T$ is the polynomial

$$a_0 + a_1 z + \cdots + a_{n-1} z^{n-1} + z^n.$$

to do

# 5C Upper-Triangular Matrices

A central goal of linear algebra is to show that given an operator $T$ on a finite-dimensional vector space $V$ , there exists a basis of $V$ with respect to which $T$ has a reasonably simple matrix.

**Theorem 5.24.** *Suppose $T \in \mathcal{L}(V)$ and $v_1, \dots, v_n$ is a basis of $V$. Then the following are equivalent.*

*(a) The matrix of $T$ with respect to $v_1, \dots, v_n$ is upper triangular.*

*(b) span$(v_1, \dots, v_k)$ is invariant under $T$ for each $k = 1, \dots, n$.*

*(c)* $Tv_k \in \operatorname{span}(v_1, \dots, v_k)$ *for each* $k = 1, \dots, n$.

**Theorem 5.25.** *Suppose* $T \in \mathcal{L}(V)$ *and* $V$ *has a basis with respect to which* $T$ *has an upper-triangular matrix with diagonal entries* $\lambda_1, \dots, \lambda_n$. *Then*

$$(T - \lambda_1 I) \cdots (T - \lambda_n I) = 0.$$

*Proof.* Let $v_1, \dots, v_n$ denote a basis of $V$ with respect to which $T$ has an upper-triangular matrix with diagonal entries $\lambda_1, \dots, \lambda_n$. Then $Tv_1 = \lambda_1 v_1$, which means that $(T - \lambda_1 I)v_1 = 0$, which implies that

$$(T - \lambda_1 I) \cdots (T - \lambda_m I)v_1 = 0$$

for $m = 1, \dots, n$ (using the commutativity of each $T - \lambda_j I$ with each $T - \lambda_k I$).

Note that $(T - \lambda_2 I)v_2 \in \operatorname{span}(v_1)$. Thus $(T - \lambda_1 I)(T - \lambda_2 I)v_2 = 0$ (by the previous paragraph), which implies that

$$(T - \lambda_1 I) \cdots (T - \lambda_m I)v_2 = 0$$

for $m = 2, \dots, n$ (using the commutativity of each $T - \lambda_j I$ with each $T - \lambda_k I$).

Continuing this pattern, we see that

$$(T - \lambda_1 I) \cdots (T - \lambda_n I)v_k = 0$$

for each $k = 1, \dots, n$. Thus $(T - \lambda_1 I) \cdots (T - \lambda_n I)$ is the 0 operator because it is 0 on each vector in a basis of $V$. ⊣

**Theorem 5.26.** *Suppose* $T \in \mathcal{L}(V)$ *has an upper-triangular matrix with respect to some basis of* $V$. *Then the eigenvalues of* $T$ *are precisely the entries on the diagonal of that upper-triangular matrix.*

*Proof.* Suppose $v_1, \dots, v_n$ is a basis of $V$ with respect to which $T$ has an upper-triangular matrix

$$\mathcal{M}(T) = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

Because $Tv_1 = \lambda_1 v_1$, we see that $\lambda_1$ is an eigenvalue of $T$.

Suppose $k \in \{2, \dots, n\}$. Then $(T - \lambda_k I)v_k \in \operatorname{span}(v_1, \dots, v_{k-1})$. Thus $T - \lambda_k I$ maps $\operatorname{span}(v_1, \dots, v_k)$ into $\operatorname{span}(v_1, \dots, v_{k-1})$. Because

$$\dim \operatorname{span}(v_1, \dots, v_k) = k \quad \text{and} \quad \dim \operatorname{span}(v_1, \dots, v_{k-1}) = k - 1,$$

this implies that $T - \lambda_k I$ restricted to $\operatorname{span}(v_1, \dots, v_k)$ is not injective. Thus there exists $v \in \operatorname{span}(v_1, \dots, v_k)$ such that $v \neq 0$ and $(T - \lambda_k I)v = 0$. Thus $\lambda_k$ is an eigenvalue of $T$. Hence we have shown that every entry on the diagonal of $\mathcal{M}(T)$ is an eigenvalue of $T$.

To prove $T$ has no other eigenvalues, let $q$ be the polynomial defined by

$$q(z) = (z - \lambda_1) \cdots (z - \lambda_n).$$

Then $q(T) = 0$. Hence $q$ is a polynomial multiple of the minimal polynomial of $T$. Thus every zero of the minimal polynomial of $T$ is a zero of $q$. Because the zeros of the minimal polynomial of $T$ are the eigenvalues of $T$, this implies that every eigenvalue of $T$ is a zero of $q$. Hence the eigenvalues of $T$ are all contained in the list $\lambda_1, \dots, \lambda_n$. ⊣

**Theorem 5.27.** *Suppose $V$ is finite-dimensional and $T \in \mathcal{L}(V)$. Then $T$ has an upper-triangular matrix with respect to some basis of $V$ if and only if the minimal polynomial of $T$ equals*

$$(z - \lambda_1) \cdots (z - \lambda_m)$$

*for some $\lambda_1, \dots, \lambda_m \in \mathbb{F}$.*

*Proof.* First suppose $T$ has an upper-triangular matrix with respect to some basis of $V$. Let $\alpha_1, \dots, \alpha_n$ denote the diagonal entries of that matrix. Define a polynomial $q \in \mathcal{P}(\mathbb{F})$ by

$$q(z) = (z - \alpha_1) \cdots (z - \alpha_n).$$

Then $q(T) = 0$ by 5.25. Hence by 5.19 $q$ is a polynomial multiple of the minimal polynomial of $T$. Thus the minimal polynomial of $T$ equals $(z - \lambda_1) \cdots (z - \lambda_m)$ for some $\lambda_1, \dots, \lambda_m \in \mathbb{F}$ with $\{\lambda_1, \dots, \lambda_m\} \subseteq \{\alpha_1, \dots, \alpha_n\}$.

To prove the implication in the other direction, now suppose the minimal polynomial of $T$ equals $(z - \lambda_1) \cdots (z - \lambda_m)$ for some $\lambda_1, \dots, \lambda_m \in \mathbb{F}$. Use induction on $m$. To get started, if $m = 1$ then $z - \lambda_1$ is the minimal polynomial of $T$, which implies that $T = \lambda_1 I$, which implies that the matrix of $T$ (with respect to any basis of $V$) is upper triangular.

Now suppose $m > 1$ and the desired result holds for all smaller positive integers. Let

$$U = \operatorname{range}(T - \lambda_m I).$$

Then $U$ is invariant under $T$. Thus $T|_U$ is an operator on $U$.

If $u \in U$, then $u = (T - \lambda_m I)v$ for some $v \in V$ and

$$(T - \lambda_1 I) \cdots (T - \lambda_{m-1} I)u = (T - \lambda_1 I) \cdots (T - \lambda_m I)v = 0.$$

Hence $(z-\lambda_1)\cdots(z-\lambda_{m-1})$ is a polynomial multiple of the minimal polynomial of $T|_U$. Thus the minimal polynomial of $T|_U$ is the product of at most $m-1$ terms of the form $z-\lambda_k$.

By our induction hypothesis, there is a basis $u_1,\dots,u_M$ of $U$ with respect to which $T|_U$ has an upper-triangular matrix. Thus for each $k \in \{1,\dots,M\}$, by 5.24 we have

$$Tu_k = (T|_U)(u_k) \in \operatorname{span}(u_1,\dots,u_k). \tag{$*$}$$

Extend $u_1,\dots,u_M$ to a basis $u_1,\dots,u_M,v_1,\dots,v_N$ of $V$. For each $k \in \{1,\dots,N\}$,

$$Tv_k = (T-\lambda_m I)v_k + \lambda_m v_k.$$

The definition of $U$ shows that $(T-\lambda_m I)v_k \in U = \operatorname{span}(u_1,\dots,u_M)$. Thus the equation above shows that

$$Tv_k \in \operatorname{span}(u_1,\dots,u_M,v_1,\dots,v_k). \tag{$**$}$$

From $(*)$ and $(**)$, we conclude by 5.24 that $T$ has an upper-triangular matrix with respect to the basis $u_1,\dots,u_M,v_1,\dots,v_N$ of $V$, as desired. ⊣

**Corollary 5.28.** *Suppose $V$ is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Then $T$ has an upper-triangular matrix with respect to some basis of $V$.*

## Exercises

**Exercise 2.** Suppose $A$ and $B$ are upper−triangular matrices of the same size, with $\alpha_1,\dots,\alpha_n$ on the diagonal of $A$ and $\beta_1,\dots,\beta_n$ on the diagonal of $B$.

(a) Show that $A+B$ is an upper−triangular matrix with $\alpha_1+\beta_1,\dots,\alpha_n+\beta_n$ on the diagonal.

(b) Show that $AB$ is an upper−triangular matrix with $\alpha_1\beta_1,\dots,\alpha_n\beta_n$ on the diagonal.

to do

# 5D Diagonalizable Operators

**Definition 5.29.** An operator on $V$ is called *diagonalizable* if the operator has a diagonal matrix with respect to some basis of $V$.

**Definition 5.30.** Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. The *eigenspace* of $T$ corresponding to $\lambda$ is the subspace $E(\lambda, T)$ of $V$ defined by

$$E(\lambda, T) = (T - \lambda I) = \{v \in V \,:\, Tv = \lambda v\}.$$

Hence $E(\lambda, T)$ is the set of all eigenvectors of $T$ corresponding to $\lambda$, along with the 0 vector.

**Theorem 5.31.** *Suppose $T \in \mathcal{L}(V)$ and $\lambda_1, \dots, \lambda_m$ are distinct eigenvalues of $T$. Then*

$$E(\lambda_1, T) + \cdots + E(\lambda_m, T)$$

*is a direct sum. Furthermore, if $V$ is finite-dimensional, then*

$$\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) \leq \dim V.$$

**Theorem 5.32.** *Suppose $V$ is finite-dimensional and $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ denote the distinct eigenvalues of $T$. Then the following are equivalent.*

*(a)* *$T$ is diagonalizable.*

*(b)* *$V$ has a basis consisting of eigenvectors of $T$.*

*(c)* $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$.

*(d)* $\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T)$.

*Proof.* We prove only (d)⇔(b). Suppose (d) holds; thus

$$\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T).$$

Choose a basis of each $E(\lambda_k, T)$; put all these bases together to form a list $v_1, \dots, v_n$ of eigenvectors of $T$, where $n = \dim V$. To show that this list is linearly independent, suppose

$$a_1 v_1 + \cdots + a_n v_n = 0,$$

where $a_1, \dots, a_n \in \mathbb{F}$. For each $k = 1, \dots, m$, let $u_k$ denote the sum of all the terms $a_j v_j$ such that $v_j \in E(\lambda_k, T)$. Thus each $u_k$ is in $E(\lambda_k, T)$, and

$$u_1 + \cdots + u_m = 0.$$

By 5.7 each $u_k$ equals 0. Thus all $a_j$'s equal 0. Thus $v_1, \dots, v_n$ is linearly independent and hence is a basis of $V$, as desired. ⊣

**Corollary 5.33.** *Suppose $V$ is finite-dimensional and $T \in \mathcal{L}(V)$ has* $\dim V$ *distinct eigenvalues. Then $T$ is diagonalizable.*

**Theorem 5.34.** *Suppose $V$ is finite-dimensional and $T \in \mathcal{L}(V)$. Then $T$ is diagonalizable if and only if the minimal polynomial of $T$ equals*

$$(z - \lambda_1) \cdots (z - \lambda_m)$$

*for some list of distinct numbers $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$.*

*Proof.* First suppose $T$ is diagonalizable. Thus there is a basis $v_1, \ldots, v_n$ of $V$ consisting of eigenvectors of $T$. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of $T$. Then for each $v_j$, there exists $\lambda_k$ with $(T - \lambda_k I)v_j = 0$. Thus

$$(T - \lambda_1 I) \cdots (T - \lambda_m I)v_j = 0,$$

which implies that the minimal polynomial of $T$ equals $(z - \lambda_1) \cdots (z - \lambda_m)$. (Some eigenvalues may be left off.)

To prove the other direction, suppose the minimal polynomial of $T$ equals $(z - \lambda_1) \cdots (z - \lambda_m)$ for some list of distinct numbers $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$. Thus

$$(T - \lambda_1 I) \cdots (T - \lambda_m I) = 0. \tag{$*$}$$

Use induction on $m$. Base case: suppose $m = 1$. Then $T - \lambda_1 I = 0$, which means that $T$ is a scalar multiple of the identity operator, which is diagonalizable.

Suppose that $m > 1$ and the desired result holds for all smaller values of $m$. The subspace $\operatorname{range}(T - \lambda_m I)$ is invariant under $T$. Thus $T|_{\operatorname{range}(T - \lambda_m I)}$ is an operator on $\operatorname{range}(T - \lambda_m I)$.

If $u \in \operatorname{range}(T - \lambda_m I)$, then $u = (T - \lambda_m I)v$ for some $v \in V$, and $(*)$ implies

$$(T - \lambda_1 I) \cdots (T - \lambda_{m-1} I)u = (T - \lambda_1 I) \cdots (T - \lambda_m I)v = 0. \tag{$**$}$$

Hence $(z - \lambda_1) \cdots (z - \lambda_{m-1})$ is a polynomial multiple of the minimal polynomial of $T$ restricted to $\operatorname{range}(T - \lambda_m I)$. Thus by our induction hypothesis, there is a basis of $\operatorname{range}(T - \lambda_m I)$ consisting of eigenvectors of $T$.

Suppose that $u \in \operatorname{range}(T - \lambda_m I) \cap \operatorname{null}(T - \lambda_m I)$. Then $Tu = \lambda_m u$. Now $(**)$ implies that

$$0 = (T - \lambda_1 I) \cdots (T - \lambda_{m-1} I)u = (\lambda_m - \lambda_1) \cdots (\lambda_m - \lambda_{m-1})u.$$

Because $\lambda_1, \ldots, \lambda_m$ are distinct, the equation above implies that $u = 0$. Hence

$$\operatorname{range}(T - \lambda_m I) \cap \operatorname{null}(T - \lambda_m I) = \{0\}.$$

Thus $\operatorname{range}(T - \lambda_m I) + \operatorname{null}(T - \lambda_m I)$ is a direct sum whose dimension is $\dim V$.

$$\operatorname{range}(T - \lambda_m I) \oplus \operatorname{null}(T - \lambda_m I) = V.$$

Every vector in $\operatorname{null}(T - \lambda_m I)$ is an eigenvector of $T$ with eigenvalue $\lambda_m$. Earlier in this proof we saw that there is a basis of $\operatorname{range}(T - \lambda_m I)$ consisting of eigenvectors of $T$. Adjoining to that basis a basis of $\operatorname{null}(T - \lambda_m I)$ gives a basis of $V$ consisting of eigenvectors of $T$, as desired. ⊣

Cf. 5.27. The next result is by 5.20.

**Corollary 5.35.** *Suppose $T \in \mathcal{L}(V)$ is diagonalizable and $U$ is a subspace of $V$ that is invariant under $T$. Then $T|_U$ is a diagonalizable operator on $U$.*

**Definition 5.36.** Suppose $T \in \mathcal{L}(V)$ and $v_1, \dots, v_n$ is a basis of $V$. Let $A$ denote the matrix of $T$ with respect to this basis. A *Gershgorin disk* of $T$ with respect to the basis $v_1, \dots, v_n$ is a set of the form

$$\left\{ z \in \mathbb{F} : |z - A_{j,j}| \leq \sum_{\substack{k=1 \\ k \neq j}}^{n} |A_{j,k}| \right\},$$

where $j \in \{1, \dots, n\}$.

**Theorem 5.37.** *Suppose $T \in \mathcal{L}(V)$ and $v_1, \dots, v_n$ is a basis of $V$. Each eigenvalue of $T$ is contained in some Gershgorin disk of $T$ with respect to the basis $v_1, \dots, v_n$.*

*Proof.* Suppose $\lambda \in \mathbb{F}$ is an eigenvalue of $T$. Let $w \in V$ be a corresponding eigenvector. There exist $c_1, \dots, c_n \in \mathbb{F}$ such that

$$w = c_1 v_1 + \cdots + c_n v_n. \tag{$*$}$$

Let $A$ denote the matrix of $T$ with respect to the basis $v_1, \dots, v_n$. Applying $T$ to both sides of the equation above gives

$$\begin{aligned} \lambda w &= \sum_{k=1}^{n} c_k T v_k \\ &= \sum_{k=1}^{n} c_k \sum_{j=1}^{n} A_{j,k} v_j \\ &= \sum_{j=1}^{n} \left( \sum_{k=1}^{n} A_{j,k} c_k \right) v_j. \end{aligned} \tag{$**$}$$

Let $j \in \{1, \dots, n\}$ be such that

$$|c_j| = \max\{|c_1|, \dots, |c_n|\}.$$

By ($*$) and ($**$)

$$\lambda c_j = \sum_{k=1}^{n} A_{j,k} c_k.$$

Subtract $A_{j,j} c_j$ from each side of the equation above and then divide both sides by $c_j$ to get

$$|\lambda - A_{j,j}| = \left| \sum_{\substack{k=1 \\ k \neq j}}^{n} A_{j,k} \frac{c_k}{c_j} \right| \leq \sum_{\substack{k=1 \\ k \neq j}}^{n} |A_{j,k}|.$$

Thus $\lambda$ is in the $j^{\text{th}}$ Gershgorin disk with respect to the basis $v_1, \dots, v_n$. ⊣

## Exercises

**Exercise 21.** The *Fibonacci sequence* $F_0, F_1, F_2, \dots$ is defined by

$$F_0 = 0, \quad F_1 = 1, \quad \text{and} \quad F_n = F_{n-2} + F_{n-1} \quad \text{for } n \geq 2.$$

Define $T \in \mathcal{L}(\mathbb{R}^2)$ by $T(x, y) = (y, x + y)$.

(a) Show that $T^n(0, 1) = (F_n, F_{n+1})$ for each nonnegative integer $n$.

(b) Find the eigenvalues of $T$.

(c) Find a basis of $\mathbb{R}^2$ consisting of eigenvectors of $T$.

(d) Use the solution to (c) to compute $T^n(0, 1)$. Conclude that

$$F_n = \frac{1}{\sqrt{5}} \left[ \left( \frac{1 + \sqrt{5}}{2} \right)^n - \left( \frac{1 - \sqrt{5}}{2} \right)^n \right]$$

for each nonnegative integer $n$.

(e) Use (d) to conclude that if $n$ is a nonnegative integer, then the Fibonacci number $F_n$ is the integer that is closest to

$$\frac{1}{\sqrt{5}} \left( \frac{1 + \sqrt{5}}{2} \right)^n.$$

to do

**Exercise 22.** Question to fill in.

to do

CHAPTER 6

# Inner Product Spaces

## 6A Inner Products and Norms

**Definition 6.1.** For $x, y \in \mathbb{R}^n$, the *dot product* of $x$ and $y$, denoted by $x \cdot y$, is defined by

$$x \cdot y = x_1 y_1 + \cdots + x_n y_n,$$

where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$.

**Definition 6.2.** An *inner product* on $V$ is a function $(u, v) \in V^2 \mapsto \langle u, v \rangle \in \mathbb{F}$ and has the following properties.

1. positivity
$$\langle v, v \rangle \geq 0 \quad \text{for all } v \in V.$$
2. definiteness
$$\langle v, v \rangle = 0 \quad \text{if and only if } v = 0.$$
3. additivity in first slot
$$\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle \quad \text{for all } u, v, w \in V.$$
4. homogeneity in first slot
$$\langle \lambda u, v \rangle = \lambda \langle u, v \rangle \quad \text{for all } \lambda \in \mathbb{F} \text{ and all } u, v \in V.$$
5. conjugate symmetry
$$\langle u, v \rangle = \overline{\langle v, u \rangle} \quad \text{for all } u, v \in V.$$

For $\lambda \in \mathbb{C}$, the notation $\lambda \geq 0$ means that $\lambda$ is real and nonnegative.

**Definition 6.3.** An inner product space is a vector space $V$ along with an inner product on $V$.

**Theorem 6.4.** *Basic properties of an inner product.*

(a) *For each fixed $v \in V$, the function that takes $u \in V$ to $\langle u, v\rangle$ is a linear map from $V$ to $\mathbb{F}$.*
(b) $\langle 0, v\rangle = 0$ *for every $v \in V$.*
(c) $\langle v, 0\rangle = 0$ *for every $v \in V$.*
(d) $\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle \quad$ *for all $u, v, w \in V$.*
(e) $\langle u, \lambda v\rangle = \overline{\lambda}\langle u, v\rangle \quad$ *for all $\lambda \in \mathbb{F}$ and all $u, v \in V$.*

**Definition 6.5.** For $v \in V$, the *norm* of $v$, denoted by $\|v\|$, is defined by

$$\|v\| = \sqrt{\langle v, v\rangle}.$$

We mainly use norms induced by inner products in this note. Cf. 6.13.

**Remark 6.6.** Suppose $v \in V$.

(a) $\|v\| = 0$ if and only if $v = 0$.
(b) $\|\lambda v\| = |\lambda|\,\|v\| \quad$ for all $\lambda \in \mathbb{F}$.

**Definition 6.7.** Say that $u, v \in V$ are *orthogonal* or $u$ is *orthogonal to* $v$ if $\langle u, v\rangle = 0$.

**Theorem 6.8.** ***Pythagorean*** *Suppose $u, v \in V$. If $u$ and $v$ are orthogonal, then*

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

**Remark 6.9.** Suppose $u, v \in V$, with $v \neq 0$. Set

$$c = \frac{\langle u, v\rangle}{\|v\|^2} \quad \text{and} \quad w = u - \frac{\langle u, v\rangle}{\|v\|^2}v.$$

Then

$$u = cv + w \quad \text{and} \quad \langle w, v\rangle = 0.$$

**Theorem 6.10.** ***Cauchy-Schwarz Inequality*** *Suppose $u, v \in V$. Then*

$$|\langle u, v\rangle| \leq \|u\| \, \|v\|.$$

*This inequality is an equality if and only if one of $u, v$ is a scalar multiple of the other.*

*Proof.* Trivial by 6.9. ⊣

**Theorem 6.11.** ***Triangle Inequality*** *Suppose $u, v \in V$. Then*

$$\|u + v\| \leq \|u\| + \|v\|.$$

*This inequality is an equality if and only if one of $u, v$ is a nonnegative real multiple of the other.*

*Proof.* Trivial by 6.10. ⊣

**Theorem 6.12.** ***Parallelogram Equality*** *Suppose $u, v \in V$. Then*

$$\|u + v\|^2 + \|u - v\|^2 = 2(\|u\|^2 + \|v\|^2).$$

## Exercises

**Exercise 15.** Suppose $u, v$ are nonzero vectors in $\mathbb{R}^2$. Prove that

$$\langle u, v\rangle = \|u\| \, \|v\| \cos\theta,$$

where $\theta$ is the angle between $u$ and $v$ (thinking of $u$ and $v$ as arrows with initial point at the origin).

Trivial. Included for that this is widely applied.

**Definition 6.13.** A *norm* on a vector space $U$ is a function $\|\cdot\| : U \to [0, \infty)$ such that the following properties hold for all $u, v \in U$ and all $\alpha \in \mathbb{F}$:

(a) $\|u\| = 0$ if and only if $u = 0$;

(b) $\|\alpha u\| = |\alpha| \|u\|$;

(c) $\|u+v\| \le \|u\| + \|v\|$.

**Exercise 28.** Prove that a norm satisfying 6.12 comes from an inner product.

*Proof.* Real case. Define
$$\langle x, y\rangle := \tfrac{1}{4}\big(\|x+y\|^2 - \|x-y\|^2\big).$$
Complex case. Define
$$\langle x, y\rangle := \tfrac{1}{4}\big(\|x+y\|^2 - \|x-y\|^2 + i\|x+iy\|^2 - i\|x-iy\|^2\big).$$

We prove only that the function in complex case is indeed an inner product. First note that $\langle x, x\rangle = \|x\|^2$ satisfies positivity and definiteness. For conjugate symmetry, note that
$$\begin{aligned}\overline{\langle x, y\rangle} &= \tfrac{1}{4}\big(\|x+y\|^2 - \|x-y\|^2 - i\|x+iy\|^2 + i\|x-iy\|^2\big)\\ &= \tfrac{1}{4}\big(\|x+y\|^2 - \|x-y\|^2 - i\,|i|\,\|-ix+y\|^2 + i\,|-i|\,\|+ix+y\|^2\big)\\ &= \tfrac{1}{4}\big(\|x+y\|^2 - \|x-y\|^2 - i\|y+ix\|^2 + i\|y-ix\|^2\big) = \langle y, x\rangle.\end{aligned}$$

Before we prove additivity and homogeneity in the first argument, set $q(x) := \|x\|^2$. The parallelogram identity is
$$q(u+v) + q(u-v) = 2q(u) + 2q(v) \qquad (\forall u, v \in V). \tag{P}$$
And define for $x, y \in V$
$$\langle x, y\rangle := \frac{1}{4}\Big(q(x+y) - q(x-y) + i\,q(x+iy) - i\,q(x-iy)\Big). \tag{C0}$$

Additivity. Fix $y \in V$ and $x_1, x_2 \in V$. First handle the real part. Apply (P) with $u = x_1 + y, v = x_2$:
$$q(x_1 + x_2 + y) + q(x_1 - x_2 + y) = 2q(x_1 + y) + 2q(x_2). \tag{A1}$$
Apply (P) with $u = x_1 - y, v = x_2$:
$$q(x_1 + x_2 - y) + q(x_1 - x_2 - y) = 2q(x_1 - y) + 2q(x_2). \tag{A2}$$
Subtract (A2) from (A1):
$$\begin{aligned}&\big(q(x_1 + x_2 + y) - q(x_1 + x_2 - y)\big) + \big(q(x_1 - x_2 + y) - q(x_1 - x_2 - y)\big)\\ &= 2\big(q(x_1 + y) - q(x_1 - y)\big).\end{aligned} \tag{A3}$$

Now swap $x_1, x_2$ in the same derivation. Using $u = x_2 + y, v = x_1$ and $u = x_2 - y, v = x_1$ we obtain

$$\begin{aligned}&\big(q(x_1 + x_2 + y) - q(x_1 + x_2 - y)\big) + \big(q(-x_1 + x_2 + y) - q(-x_1 + x_2 - y)\big)\\ &= 2\big(q(x_2 + y) - q(x_2 - y)\big).\end{aligned} \tag{A4}$$

Since $q$ is even, $q(-x_1 + x_2 \pm y) = q(x_1 - x_2 \mp y)$. Substitute these into (A4) to get

$$\begin{aligned}&\big(q(x_1 + x_2 + y) - q(x_1 + x_2 - y)\big) + \big(q(x_1 - x_2 - y) - q(x_1 - x_2 + y)\big)\\ &= 2\big(q(x_2 + y) - q(x_2 - y)\big).\end{aligned} \tag{A5}$$

Adding (A3) and (A5) cancels the middle terms and yields

$$\begin{aligned}&2\big(q(x_1 + x_2 + y) - q(x_1 + x_2 - y)\big)\\ &= 2\Big(\big(q(x_1 + y) - q(x_1 - y)\big) + \big(q(x_2 + y) - q(x_2 - y)\big)\Big).\end{aligned} \tag{A6}$$

Divide by 8 to obtain

$$\operatorname{Re}\langle x_1 + x_2, y\rangle = \operatorname{Re}\langle x_1, y\rangle + \operatorname{Re}\langle x_2, y\rangle. \tag{A7}$$

For the imaginary part, replace $y$ by $iy$ in the previous argument. Repeating (A1)–(A6) with $y \mapsto iy$ gives

$$\operatorname{Re}\langle x_1 + x_2, iy\rangle = \operatorname{Re}\langle x_1, iy\rangle + \operatorname{Re}\langle x_2, iy\rangle. \tag{A8}$$

But by (C0),

$$\operatorname{Im}\langle x, y\rangle = \frac{1}{4}\big(q(x + iy) - q(x - iy)\big) = \operatorname{Re}\langle x, iy\rangle. \tag{A9}$$

Combining (A8) and (A9) gives

$$\operatorname{Im}\langle x_1 + x_2, y\rangle = \operatorname{Im}\langle x_1, y\rangle + \operatorname{Im}\langle x_2, y\rangle. \tag{A10}$$

Together, (A7) and (A10) imply

$$\langle x_1 + x_2, y\rangle = \langle x_1, y\rangle + \langle x_2, y\rangle. \tag{Add}$$

Homogeneity.

(a) Integer and rational real scalars. From (Add) with $x_1 = x_2 = \cdots = x$ (induction) and the identity $\langle -x, y\rangle = -\langle x, y\rangle$ (obvious from (C0)), we have

$$\langle nx, y\rangle = n\langle x, y\rangle \quad (\forall n \in \mathbb{Z}). \tag{H1}$$

Let $r = \frac{m}{n} \in \mathbb{Q}$ with $n \neq 0$. Then by (H1),

$$n\langle rx, y\rangle = \langle nrx, y\rangle = \langle mx, y\rangle = m\langle x, y\rangle,$$

hence

$$\langle rx, y\rangle = r\langle x, y\rangle \quad (\forall r \in \mathbb{Q}). \tag{H2}$$

(b) Real scalars by continuity. The map $t \mapsto \langle tx, y\rangle$ is continuous because it is a linear combination of continuous maps $t \mapsto q(tx \pm y)$. Since $\mathbb{Q}$ is dense, (H2) extends to all $t \in \mathbb{R}$:

$$\langle tx, y\rangle = t\langle x, y\rangle \quad (\forall t \in \mathbb{R}). \tag{H3}$$

(c) Multiplication by $i$. Using (C0) and the norm homogeneity $\|iz\| = \|z\|$,

$$\begin{aligned}\langle ix, y\rangle &= \frac{1}{4}\Big(q(ix+y) - q(ix-y) + i\,q(ix+iy) - i\,q(ix-iy)\Big)\\ &= \frac{1}{4}\Big(q(x-iy) - q(x+iy) + i\,q(x+y) - i\,q(x-y)\Big)\\ &= -\frac{1}{4}\big(q(x+iy) - q(x-iy)\big) \;+\; i\,\frac{1}{4}\big(q(x+y) - q(x-y)\big)\\ &= -\operatorname{Im}\langle x, y\rangle \;+\; i\operatorname{Re}\langle x, y\rangle \;=\; i\big(\operatorname{Re}\langle x, y\rangle + i\operatorname{Im}\langle x, y\rangle\big) \;=\; i\langle x, y\rangle.\end{aligned} \tag{H4}$$

(d) General complex scalars. For $\alpha = a + ib$ with $a, b \in \mathbb{R}$, using (H3), (H4), and (Add) with $x_1 = ax$, $x_2 = bx$,

$$\begin{aligned}\langle \alpha x, y\rangle &= \langle ax + b(ix), y\rangle = \langle ax, y\rangle + \langle b(ix), y\rangle = a\langle x, y\rangle + b\,\langle ix, y\rangle\\ &= (a+ib)\langle x, y\rangle = \alpha\langle x, y\rangle.\end{aligned} \tag{Hom}$$

(Add) and (Hom) gives linearity of $\langle \cdot, y\rangle$ in the first argument over $\mathbb{C}$. ⊣

## 6B Orthonormal Bases

**Definition 6.14.** A list of vectors is called *orthonormal* if each vector in the list has norm 1 and is orthogonal to all the other vectors in the list.

**Theorem 6.15.** *If $e_1, \dots, e_m$ is an orthonormal list of vectors in $V$, then*

$$\left\|a_1 e_1 + \cdots + a_m e_m\right\|^2 = |a_1|^2 + \cdots + |a_m|^2$$

*for all $a_1, \dots, a_m \in \mathbb{F}$.*

**Corollary 6.16.** *Every orthonormal list of vectors is linearly independent.*

**Theorem 6.17.** ***Bessel's Inequality*** *Suppose $e_1, \dots, e_m$ is an orthonormal list of vectors in $V$. If $v \in V$ then*

$$|\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_m\rangle|^2 \leq \|v\|^2.$$

*Proof.* Suppose $v \in V$. Then

$$v = \underbrace{\langle v, e_1\rangle e_1 + \cdots + \langle v, e_m\rangle e_m}_{u} + \underbrace{v - \langle v, e_1\rangle e_1 - \cdots - \langle v, e_m\rangle e_m}_{w}.$$

Let $u$ and $w$ be defined as in the equation above. If $k \in \{1, \dots, m\}$, then

$$\langle w, e_k\rangle = \langle v, e_k\rangle - \langle v, e_k\rangle\langle e_k, e_k\rangle = 0.$$

This implies that $\langle w, u\rangle = 0$. By 6.8 and 6.15,

$$\|v\|^2 = \|u\|^2 + \|w\|^2 \geq \|u\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_m\rangle|^2. \qquad \dashv$$

**Definition 6.18.** An *orthonormal basis* of $V$ is an orthonormal list of vectors in $V$ that is also a basis of $V$.

By 6.16 and 2.24, we have that

**Theorem 6.19.** *Suppose $V$ is finite-dimensional. Then every orthonormal list of vectors in $V$ of length $\dim V$ is an orthonormal basis of $V$.*

**Theorem 6.20.** *Suppose $e_1, \dots, e_n$ is an orthonormal basis of $V$ and $u, v \in V$. Then*

*(a)* $v = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n$,
*(b)* $\|v\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2$,
*(c)* $\langle u, v\rangle = \langle u, e_1\rangle\overline{\langle v, e_1\rangle} + \cdots + \langle u, e_n\rangle\overline{\langle v, e_n\rangle}$.

Note that each inner product space of dimension $n$ (when given an orthonormal basis) behaves like $\mathbb{F}^n$, with the role of the coordinates of a vector in $\mathbb{F}^n$ played by $\langle v, e_1\rangle, \dots, \langle v, e_n\rangle$.

**Theorem 6.21.** ***Gram–Schmidt Procedure*** *Suppose $v_1, \dots, v_m$ is a linearly independent list* *of vectors in $V$. Let $f_1 = v_1$. For $k = 2, \dots, m$, define $f_k$ inductively by*

$$f_k = v_k - \frac{\langle v_k, f_1 \rangle}{\|f_1\|^2} f_1 - \cdots - \frac{\langle v_k, f_{k-1} \rangle}{\|f_{k-1}\|^2} f_{k-1}.$$

*For each $k = 1, \dots, m$, let*

$$e_k = \frac{f_k}{\|f_k\|}.$$

*Then $e_1, \dots, e_m$ is an orthonormal list of vectors in $V$ such that*

$$\operatorname{span}(v_1, \dots, v_k) = \operatorname{span}(e_1, \dots, e_k)$$

*for each $k = 1, \dots, m$.*

**Corollary 6.22.** *Every finite-dimensional inner product space has an orthonormal basis.*

**Corollary 6.23.** *Suppose $V$ is finite-dimensional. Then every orthonormal list of vectors in $V$ can be extended to an orthonormal basis of $V$.*

**Theorem 6.24.** *Suppose $V$ is finite-dimensional and $T \in \mathcal{L}(V)$. Then $T$ has an upper-triangular matrix with respect to some orthonormal basis of $V$ if and only if the minimal polynomial of $T$ equals*

$$(z - \lambda_1) \cdots (z - \lambda_m)$$

*for some $\lambda_1, \dots, \lambda_m \in \mathbb{F}$.*

*Proof.* Suppose $T$ has an upper-triangular matrix with respect to some basis $v_1, \dots, v_n$ of $V$. Thus $\operatorname{span}(v_1, \dots, v_k)$ is invariant under $T$ for each $k = 1, \dots, n$ by 5.24. Apply the Gram–Schmidt procedure to $v_1, \dots, v_n$, producing an orthonormal basis $e_1, \dots, e_n$ of $V$. Because

$$\operatorname{span}(e_1, \dots, e_k) = \operatorname{span}(v_1, \dots, v_k)$$

for each $k$ by 6.21, we conclude that $\operatorname{span}(e_1, \dots, e_k)$ is invariant under $T$ for each $k = 1, \dots, n$. Thus, by 5.24, $T$ has an upper-triangular matrix with respect to the orthonormal basis $e_1, \dots, e_n$. Now 5.27 completes the proof. ⊣

**Corollary 6.25.** ***Schur*** *Every operator on a finite-dimensional complex inner product space has an upper-triangular matrix with respect to some orthonormal basis.*

**Theorem 6.26.** ***Riesz*** *Suppose $V$ is finite-dimensional and $\varphi$ is a linear functional on $V$. Then there is a unique vector $v \in V$ such that*

$$\varphi(u) = \langle u, v \rangle$$

*for every $u \in V$.*

*Proof.* First we show the existence of $v$. Let $e_1, \dots, e_n$ be an orthonormal basis of $V$. Then

$$\begin{aligned} \varphi(u) &= \varphi\big(\langle u, e_1 \rangle e_1 + \cdots + \langle u, e_n \rangle e_n\big) \\ &= \langle u, e_1 \rangle \varphi(e_1) + \cdots + \langle u, e_n \rangle \varphi(e_n) \\ &= \Big\langle u, \overline{\varphi(e_1)}\, e_1 + \cdots + \overline{\varphi(e_n)}\, e_n \Big\rangle \end{aligned}$$

for every $u \in V$. Thus setting

$$v = \overline{\varphi(e_1)}\, e_1 + \cdots + \overline{\varphi(e_n)}\, e_n,$$

we have $\varphi(u) = \langle u, v \rangle$ for every $u \in V$.

Now suppose $v_1, v_2 \in V$ are such that

$$\varphi(u) = \langle u, v_1 \rangle = \langle u, v_2 \rangle$$

for every $u \in V$. Then

$$0 = \langle u, v_1 \rangle - \langle u, v_2 \rangle = \langle u, v_1 - v_2 \rangle$$

for every $u \in V$. Taking $u = v_1 - v_2$ gives

$$\|v_1 - v_2\|^2 = 0,$$

which implies $v_1 = v_2$, as desired. ⊣

## Exercises

**Exercise 10.** Suppose $v_1, \dots, v_m$ is a linearly independent list in $V$. Explain why the orthonormal list produced by the formulas of the Gram–Schmidt procedure 6.21 is the only orthonormal list $e_1, \dots, e_m$ in $V$ such that $\langle v_k, e_k \rangle > 0$ and

$$\operatorname{span}(v_1, \dots, v_k) = \operatorname{span}(e_1, \dots, e_k)$$

for each $k = 1, \ldots, m$.

to do

## 6C Orthogonal Complements and Minimization Problems

**Definition 6.27.** If $U$ is a subset of $V$, then the *orthogonal complement* of $U$, denoted by $U^\perp$, is the set of all vectors in $V$ that are orthogonal to every vector in $U$:

$$U^\perp = \{ v \in V : \langle u, v \rangle = 0 \text{ for every } u \in U \}.$$

**Theorem 6.28.** *Properties of orthogonal complement.*

*(a) If $U$ is a subset of $V$, then $U^\perp$ is a subspace of $V$.*

*(b) $\{0\}^\perp = V$.*

*(c) $V^\perp = \{0\}$.*

*(d) If $U$ is a subset of $V$, then $U \cap U^\perp \subseteq \{0\}$.*

*(e) If $G$ and $H$ are subsets of $V$ and $G \subseteq H$, then $H^\perp \subseteq G^\perp$.*

**Theorem 6.29.** *Suppose $U$ is a finite-dimensional subspace of $V$. Then*

$$V = U \oplus U^\perp.$$

The proof is trivial, if we consider 1.18 and an approach similar to 6.17.

**Corollary 6.30.** *Suppose $V$ is finite-dimensional and $U$ is a subspace of $V$. Then*

$$\dim U^\perp = \dim V - \dim U.$$

**Corollary 6.31.** *Suppose $U$ is a finite-dimensional subspace of $V$. Then*

$$U = (U^\perp)^\perp.$$

**Corollary 6.32.** *Suppose $U$ is a finite-dimensional subspace of $V$. Then*

$$U^\perp = \{0\} \iff U = V.$$

**Definition 6.33.** Suppose $U$ is a finite-dimensional subspace of $V$. The *orthogonal projection* of $V$ onto $U$ is the operator $P_U \in \mathcal{L}(V)$ defined as follows: For each $v \in V$, write $v = u + w$, where $u \in U$ and $w \in U^\perp$. Then let

$$P_U v = u.$$

**Theorem 6.34.** *Suppose $U$ is a finite-dimensional subspace of $V$. Then*

(a) $P_U \in \mathcal{L}(V)$;
(b) $P_U u = u$ *for every* $u \in U$;
(c) $P_U w = 0$ *for every* $w \in U^\perp$;
(d) $\operatorname{range} P_U = U$;
(e) $\operatorname{null} P_U = U^\perp$;
(f) $v - P_U v \in U^\perp$ *for every* $v \in V$;
(g) $P_U^2 = P_U$;
(h) $\|P_U v\| \leq \|v\|$ *for every* $v \in V$;
(i) *if* $e_1, \dots, e_m$ *is an orthonormal basis of* $U$ *and* $v \in V$, *then*

$$P_U v = \langle v, e_1 \rangle e_1 + \dots + \langle v, e_m \rangle e_m.$$

**Theorem 6.35.** ***Riesz*** *Suppose $V$ is finite-dimensional. For each $v \in V$, define $\varphi_v \in V'$ by*

$$\varphi_v(u) = \langle u, v \rangle$$

*for each $u \in V$. Then $v \mapsto \varphi_v$ is a one-to-one function from $V$ onto $V'$.*

*Proof.* Cf. 6.26. We prove only the surjectivity.

Let $\varphi \in V'$ be given. If $\varphi = 0$, then $\varphi = \varphi_0$. Thus assume $\varphi \neq 0$. Then $\operatorname{null} \varphi \neq V$, which implies

$$(\operatorname{null} \varphi)^\perp \neq \{0\}.$$

Let $w \in (\operatorname{null} \varphi)^\perp$ with $w \neq 0$ and define

$$v = \frac{\overline{\varphi(w)}}{\|w\|^2}\, w.$$

Then $v \in (\operatorname{null} \varphi)^\perp$ and $v \neq 0$ because $w \notin \operatorname{null} \varphi$. Taking the norm of both sides of $(*)$ gives

$$\|v\| = \frac{|\varphi(w)|}{\|w\|}.$$

Applying $\varphi$ to both sides of $(*)$ we get

$$\varphi(v) = \frac{|\varphi(w)|^2}{\|w\|^2} = \|v\|^2.$$

Now suppose $u \in V$. Using the equation above, we can write

$$u = \left(u - \frac{\varphi(u)}{\varphi(v)} v\right) + \frac{\varphi(u)}{\|v\|^2} v.$$

The first term in parentheses lies in $\operatorname{null} \varphi$ and is therefore orthogonal to $v$. Taking the inner product of both sides with $v$ gives

$$\langle u, v\rangle = \frac{\varphi(u)}{\|v\|^2}\langle v, v\rangle = \varphi(u).$$

Thus $\varphi = \varphi_v$, showing that $v \longmapsto \varphi_v$ is surjective, as desired. ⊣

**Theorem 6.36.** *Suppose $U$ is a finite-dimensional subspace of $V$, $v \in V$, and $u \in U$. Then*

$$\|v - P_U v\| \leq \|v - u\|.$$

*Furthermore, the inequality above is an equality if and only if $u = P_U v$.*

*Proof.* We have

$$\begin{aligned} \|v - P_U v\|^2 &\leq \|v - P_U v\|^2 + \|P_U v - u\|^2 \\ &= \|(v - P_U v) + (P_U v - u)\|^2 \\ &= \|v - u\|^2. \end{aligned}$$

⊣

This result is often combined with the formula 6.34(i) to compute explicit solutions to minimization problems.

**Theorem 6.37.** *Suppose $V$ is finite-dimensional and $T \in \mathcal{L}(V, W)$. Then*

$$T|_{(\operatorname{null} T)^\perp}$$

*is a one-to-one map of* $(\text{null } T)^{\perp}$ *onto* $\text{range } T$.

*Proof.* Suppose that $v \in (\text{null } T)^{\perp}$ and $T|_{(\text{null } T)^{\perp}} v = 0$. Hence $Tv = 0$ and thus $v \in (\text{null } T) \cap (\text{null } T)^{\perp}$, which implies that $v = 0$. Hence $\text{null } T|_{(\text{null } T)^{\perp}} = \{0\}$, which implies that $T|_{(\text{null } T)^{\perp}}$ is injective, as desired.

Suppose $w \in \text{range } T$. Hence there exists $v \in V$ such that $w = Tv$. There exist $u \in \text{null } T$ and $x \in (\text{null } T)^{\perp}$ such that $v = u + x$. Now

$$T|_{(\text{null } T)^{\perp}} x = Tx = Tv - Tu = w - 0 = w,$$

which shows that $w \in \text{range } T|_{(\text{null } T)^{\perp}}$. Hence $\text{range } T \subseteq \text{range } T|_{(\text{null } T)^{\perp}}$. ⊣

**Definition 6.38.** Suppose that $V$ is finite-dimensional and $T \in \mathcal{L}(V, W)$. The *pseudoinverse* $T^{\dagger} \in \mathcal{L}(W, V)$ of $T$ is the linear map from $W$ to $V$ defined by

$$T^{\dagger} w = \left(T|_{(\text{null } T)^{\perp}}\right)^{-1} P_{\text{range } T} w$$

for each $w \in W$.

**Theorem 6.39.** *Suppose $V$ is finite-dimensional and $T \in \mathcal{L}(V, W)$.*

*(a) If $T$ is invertible, then $T^{\dagger} = T^{-1}$.*
*(b) $TT^{\dagger} = P_{\text{range } T}$ = the orthogonal projection of $W$ onto* $\text{range } T$.
*(c) $T^{\dagger}T = P_{(\text{null } T)^{\perp}}$ = the orthogonal projection of $V$ onto* $(\text{null } T)^{\perp}$.

**Theorem 6.40.** *Suppose $V$ is finite-dimensional, $T \in \mathcal{L}(V, W)$, and $b \in W$.*

*(a) If $x \in V$, then*
$$\|T(T^{\dagger} b) - b\| \le \|Tx - b\|,$$
*with equality if and only if $x \in T^{\dagger} b + \text{null } T$.*

*(b) If $x \in T^{\dagger} b + \text{null } T$, then*
$$\|T^{\dagger} b\| \le \|x\|,$$
*with equality if and only if $x = T^{\dagger} b$.*

*Proof.* (a) is by 6.36 and 6.39 (b). (b) The definition of $T^{\dagger}$ implies that $T^{\dagger} b \in (\text{null } T)^{\perp}$. Thus the Pythagorean theorem implies that

$$\|T^{\dagger} b\| \le \|x\|.$$ ⊣

Chapter 7

# Operators on Inner Product Spaces

Assume from now on that $V$ and $W$ are finite dimensional vector spaces.

## 7A Self-Adjoint and Normal Operators

**Definition 7.1.** Suppose $T \in \mathcal{L}(V, W)$. The *adjoint* of $T$ is the function $T^* : W \to V$ such that
$$\langle Tv, w\rangle = \langle v, T^* w\rangle$$
for every $v \in V$ and every $w \in W$.

Consider the linear functional $v \mapsto \langle Tv, w\rangle$. By 6.26 there exists a unique vector in $V$ such that this linear functional is given by taking the inner product with it, which is called $T^* w$.

**Theorem 7.2.** *If $T \in \mathcal{L}(V, W)$, then $T^* \in \mathcal{L}(W, V)$.*

*Proof.* Suppose $T \in \mathcal{L}(V, W)$. If $v \in V$ and $w_1, w_2 \in W$, then
$$\begin{aligned}\langle Tv, w_1 + w_2\rangle &= \langle Tv, w_1\rangle + \langle Tv, w_2\rangle \\ &= \langle v, T^* w_1\rangle + \langle v, T^* w_2\rangle \\ &= \langle v, T^* w_1 + T^* w_2\rangle.\end{aligned}$$

The equation above shows that
$$T^*(w_1 + w_2) = T^* w_1 + T^* w_2.$$

If $v \in V$, $\lambda \in \mathbb{F}$, and $w \in W$, then
$$\begin{aligned}\langle Tv, \lambda w\rangle &= \overline{\lambda}\langle Tv, w\rangle \\ &= \overline{\lambda}\langle v, T^* w\rangle \\ &= \langle v, \lambda T^* w\rangle.\end{aligned}$$

The equation above shows that

$$T^*(\lambda w) = \lambda T^* w. \qquad \dashv$$

**Theorem 7.3.** *Suppose $T \in \mathcal{L}(V, W)$. Then*

*(a)* $(S + T)^* = S^* + T^*$ *for all* $S \in \mathcal{L}(V, W)$*;*
*(b)* $(\lambda T)^* = \overline{\lambda}\, T^*$ *for all* $\lambda \in \mathbb{F}$*;*
*(c)* $(T^*)^* = T$*;*
*(d)* $(ST)^* = T^* S^*$ *for all* $S \in \mathcal{L}(W, U)$ *(here $U$ is a finite-dimensional inner product space over* $\mathbb{F}$*);*
*(e)* $I^* = I$*, where $I$ is the identity operator on $V$;*
*(f) if $T$ is invertible, then $T^*$ is invertible and* $(T^*)^{-1} = (T^{-1})^*$.

*Proof.* We prove only (f). Note that

$$\langle v, w\rangle = \langle T \cdot T^{-1} v, w\rangle = \langle v, (T^{-1})^* T^* w\rangle.$$

Thus $(T^{-1})^* T^* = I$. So $T^*$ is invertible, and $(T^*)^{-1} = (T^{-1})^*$. $\dashv$

*Comment.* This proof is technically not correct. Cf. 3.33. We should instead take adjoints of both sides of the equations $TT^{-1} = I$ and $T^{-1}T = I$. Note also that if $\mathbb{F} = \mathbb{R}$, then $T \mapsto T^* \in \mathcal{L}(\mathcal{L}(V, W), \mathcal{L}(W, V))$.

**Theorem 7.4.** *Suppose $T \in \mathcal{L}(V, W)$. Then*

*(a)* null $T^* = (\text{range } T)^\perp$*;*
*(b)* range $T^* = (\text{null } T)^\perp$*;*
*(c)* null $T = (\text{range } T^*)^\perp$*;*
*(d)* range $T = (\text{null } T^*)^\perp$.

*Proof.* Let $w \in W$. Then

$$\begin{aligned} w \in \text{null } T^* &\iff T^* w = 0 \\ &\iff \langle v, T^* w\rangle = 0 \text{ for all } v \in V \\ &\iff \langle Tv, w\rangle = 0 \text{ for all } v \in V \\ &\iff w \in (\text{range } T)^\perp. \end{aligned}$$

Thus null $T^* = (\text{range } T)^\perp$, proving (a).

If we take the orthogonal complement of both sides of (a), we get (d). Replacing $T$ with $T^*$ in (a) gives (c). Finally, replacing $T$ with $T^*$ in (d) gives (b). $\dashv$

**Theorem 7.5.** *Let $T \in \mathcal{L}(V, W)$. Suppose $e_1, \dots, e_n$ is an orthonormal basis of $V$ and $f_1, \dots, f_m$ is an orthonormal basis of $W$. Then*

$$\mathcal{M}(T^*) = \big(\mathcal{M}(T)\big)^*,$$

*where the first $*$ denotes adjoint (of maps) and the second denotes conjugate transpose (of matrices).*

*Proof.* Recall that we obtain the $k$th column of $\mathcal{M}(T)$ by writing $Te_k$ as a linear combination of the $f_j$'s; the scalars used in this linear combination then become the $k$th column of $\mathcal{M}(T)$. Because $f_1, \dots, f_m$ is an orthonormal basis of $W$, we know how to write $Te_k$ as a linear combination of the $f_j$'s:

$$Te_k = \langle Te_k, f_1\rangle f_1 + \cdots + \langle Te_k, f_m\rangle f_m.$$

Thus the entry in row $j$, column $k$, of $\mathcal{M}(T)$ is $\langle Te_k, f_j\rangle$.

In the statement above, replace $T$ with $T^*$ and interchange $e_1, \dots, e_n$ and $f_1, \dots, f_m$. This shows that the entry in row $j$, column $k$, of $\mathcal{M}(T^*)$ is

$$\langle T^* f_k, e_j\rangle = \overline{\langle f_k, Te_j\rangle} = \langle Te_j, f_k\rangle,$$

which equals the complex conjugate of the entry in row $k$, column $j$, of $\mathcal{M}(T)$. Thus

$$\mathcal{M}(T^*) = \big(\mathcal{M}(T)\big)^*. \qquad \dashv$$

**Definition 7.6.** An operator $T \in \mathcal{L}(V)$ is called *self-adjoint* if $T = T^*$.

**Remark 7.7.** A good analogy to keep in mind is that the adjoint on $\mathcal{L}(V)$ plays a role similar to that of the complex conjugate on $\mathbb{C}$. A complex number $z$ is real if and only if $z = \overline{z}$; thus a self-adjoint operator ($T = T^*$) is analogous to a real number.

**Theorem 7.8.** *Every eigenvalue of a self-adjoint operator is real.*

*Proof.* Suppose $T$ is a self-adjoint operator on $V$. Let $\lambda$ be an eigenvalue of $T$, and let $v$ be a nonzero vector in $V$ such that $Tv = \lambda v$. Then

$$\lambda\|v\|^2 = \langle \lambda v, v\rangle = \langle Tv, v\rangle = \langle v, Tv\rangle = \langle v, \lambda v\rangle = \overline{\lambda}\|v\|^2.$$

Thus $\lambda = \overline{\lambda}$, which means that $\lambda$ is real, as desired. $\dashv$

**Theorem 7.9.** *Suppose $V$ is a complex inner product space and $T \in \mathcal{L}(V)$. Then*

$$\langle Tv, v\rangle = 0 \quad \text{for every } v \in V \qquad \iff \qquad T = 0.$$

*Proof.* If $u, w \in V$, then

$$\langle Tu, w\rangle = \frac{\langle T(u+w), u+w\rangle - \langle T(u-w), u-w\rangle}{4} + \frac{\langle T(u+iw), u+iw\rangle - \langle T(u-iw), u-iw\rangle}{4} i.$$

Now suppose $\langle Tv, v\rangle = 0$ for every $v \in V$. Then the equation above implies that $\langle Tu, w\rangle = 0$ for all $u, w \in V$, which then implies that $Tu = 0$ for every $u \in V$. Hence $T = 0$, as desired. ⊣

**Theorem 7.10.** *Suppose $V$ is a complex inner product space and $T \in \mathcal{L}(V)$. Then*

$$T \text{ is self-adjoint} \iff \langle Tv, v\rangle \in \mathbb{R} \quad \text{for every } v \in V.$$

*Proof.* By 7.9,

$$\begin{aligned} T \text{ is self-adjoint} &\iff T - T^* = 0 \\ &\iff \langle (T - T^*)v, v\rangle = 0 \text{ for every } v \in V \\ &\iff \langle Tv, v\rangle - \overline{\langle Tv, v\rangle} = 0 \text{ for every } v \in V \\ &\iff \langle Tv, v\rangle \in \mathbb{R} \text{ for every } v \in V. \end{aligned}$$

⊣

**Theorem 7.11.** *Suppose $T$ is a self-adjoint operator on $V$. Then*

$$\langle Tv, v\rangle = 0 \text{ for every } v \in V \iff T = 0.$$

*Proof.* By 7.9 we can assume that $V$ is a real inner product space. If $u, w \in V$, then

$$\langle Tu, w\rangle = \frac{\langle T(u+w), u+w\rangle - \langle T(u-w), u-w\rangle}{4},$$

as can be proved by computing the right side using the equation

$$\langle Tw, u\rangle = \langle w, Tu\rangle = \langle Tu, w\rangle,$$

Hence $T = 0$, as desired. ⊣

**Definition 7.12.** An operator on an inner product space is called *normal* if it commutes with its adjoint.

Note that every self-adjoint operator is normal.

**Theorem 7.13.** *Suppose $T \in \mathcal{L}(V)$. Then*

$$T \text{ is normal} \iff \|Tv\| = \|T^*v\| \text{ for every } v \in V.$$

*Proof.* By that $(T^*T - TT^*)$ is self-adjoint we have

$$\begin{aligned} T \text{ is normal} &\iff T^*T - TT^* = 0 \\ &\iff \langle (T^*T - TT^*)v, v\rangle = 0 \quad \text{for every } v \in V \\ &\iff \langle T^*Tv, v\rangle = \langle TT^*v, v\rangle \quad \text{for every } v \in V \\ &\iff \langle Tv, Tv\rangle = \langle T^*v, T^*v\rangle \quad \text{for every } v \in V \\ &\iff \|Tv\|^2 = \|T^*v\|^2 \quad \text{for every } v \in V \\ &\iff \|Tv\| = \|T^*v\| \quad \text{for every } v \in V. \end{aligned}$$

⊣

**Theorem 7.14.** *Suppose $T \in \mathcal{L}(V)$ is normal. Then*

*(a)* null $T$ = null $T^*$;

*(b)* range $T$ = range $T^*$;

*(c)* $V$ = null $T \oplus$ range $T$;

*(d)* $T - \lambda I$ *is normal for every* $\lambda \in \mathbb{F}$;

*(e) if* $v \in V$ *and* $\lambda \in \mathbb{F}$, *then* $Tv = \lambda v$ *if and only if* $T^*v = \bar{\lambda}v$.

*Proof.* For (b) and (c), note that by (a) and 7.4 (d)

$$\text{range } T = (\text{null } T^*)^\perp = (\text{null } T)^\perp.$$

By (d) and 7.13, we have that

$$\|(T - \lambda I)v\| = 0 \iff \|(T^* - \bar{\lambda} I)v\| = 0,$$

proving (e). ⊣

**Theorem 7.15.** *Suppose $T \in \mathcal{L}(V)$ is normal. Then eigenvectors of $T$ corresponding to distinct eigenvalues are orthogonal.*

*Proof.* Note that $T - \lambda I$ is normal. The rest is trivial by 7.14. $\dashv$

**Theorem 7.16.** *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Then $T$ is normal if and only if there exist commuting self-adjoint operators $A$ and $B$ such that*

$$T = A + iB.$$

*Proof.* First suppose $T$ is normal. Let

$$A = \frac{T + T^*}{2} \quad \text{and} \quad B = \frac{T - T^*}{2i}. \tag{$*$}$$

Then $A$ and $B$ are self-adjoint and $T = A + iB$. A quick computation shows that

$$AB - BA = \frac{T^*T - TT^*}{2i}. \tag{$**$}$$

Because $T$ is normal, the right side of the equation above equals 0. Thus the operators $A$ and $B$ commute, as desired.

To prove the implication in the other direction, now suppose there exist commuting self-adjoint operators $A$ and $B$ such that $T = A + iB$. Then $T^* = A - iB$. Adding the last two equations and then dividing by 2 produces the equation for $A$ in $(*)$. Subtracting the last two equations and then dividing by $2i$ produces the equation for $B$ in $(*)$. Now $(*)$ implies $(**)$. Because $B$ and $A$ commute, $(**)$ implies that $T$ is normal, as desired. $\dashv$

Cf. 7A.12. Obviously $A$ and $B$ are unique. Also note that this decomposition exists even if $T$ is not normal.

## Exercises

**Exercise 7.** Question to fill in.

to do

**Exercise 8.** Question to fill in.

to do

**Exercise 12.** An operator $B \in \mathcal{L}(V)$ is called *skew* if

$$B^* = -B.$$

Suppose that $T \in \mathcal{L}(V)$. Prove that $T$ is normal if and only if there exist commuting operators $A$ and $B$ such that $A$ is self-adjoint, $B$ is a skew operator, and

$$T = A + B.$$

*Proof.* First suppose $T$ is normal. Let

$$A = \frac{T + T^*}{2} \quad \text{and} \quad B = \frac{T - T^*}{2}. \tag{$*$}$$

Then $A$ is self-adjoint $B$ is skew. Also

$$AB - BA = \frac{T^*T - TT^*}{2}. \tag{$**$}$$

Because $T$ is normal, the right side of the equation above equals 0. Thus the operators $A$ and $B$ commute, as desired.

To prove the implication in the other direction, now suppose there exist commuting operators $A$ and $B$ such that $A$ is self-adjoint, $B$ is a skew operator and $T = A + B$. Then $T^* = A - B$. Adding the last two equations and then dividing by 2 produces the equation for $A$ in $(*)$. Subtracting the last two equations and then dividing by $2i$ produces the equation for $B$ in $(*)$. Now $(*)$ implies $(**)$. Because $B$ and $A$ commute, $(**)$ implies that $T$ is normal, as desired. ⊣

**Exercise 27.** Suppose $T \in \mathcal{L}(V)$ is normal. Prove that

$$\operatorname{null} T^k = \operatorname{null} T \quad \text{and} \quad \operatorname{range} T^k = \operatorname{range} T$$

for every positive integer $k$.

*Proof.* We may assume $k \geq 2$. We first prove that $\operatorname{null} T^k = \operatorname{null} T$. That $\operatorname{null} T \subseteq \operatorname{null} T^k$ is trivial. To prove the reverse inclusion, let $v \in \operatorname{null} T^k$. Then

$$\langle T^*T^{k-1}v,\ T^*T^{k-1}v\rangle = \langle TT^*T^{k-1}v,\ T^{k-1}v\rangle = \langle T^*T^kv,\ T^{k-1}v\rangle = 0,$$

where the second equality uses the normality of $T$. Thus

$$0 = \langle T^*T^{k-1}v, T^{k-2}v\rangle = \langle T^{k-1}v, T^{k-1}v\rangle.$$

Hence $T^{k-1}v = 0$, or equivalently $v \in \operatorname{null} T^{k-1}$. Repeating this argument with $k$ replaced by $k-1$, we obtain $v \in \operatorname{null} T^{k-2}$, and so on, until we reach $v \in \operatorname{null} T$. Thus $\operatorname{null} T^k \subseteq \operatorname{null} T$, proving

$$\operatorname{null} T^k = \operatorname{null} T.$$

Now we show that $\operatorname{range} T^k = \operatorname{range} T$. That $\operatorname{range} T^k \subseteq \operatorname{range} T$ is trivial. Note

$$\dim \operatorname{range} T^k = \dim V - \dim \operatorname{null} T^k = \dim V - \dim \operatorname{null} T = \dim \operatorname{range} T.$$

Hence $\operatorname{range} T^k = \operatorname{range} T$. This completes the proof. $\dashv$

## 7B Spectral Theorem

**Theorem 7.17.** *Suppose $T \in \mathcal{L}(V)$ is self-adjoint and $b, c \in \mathbb{R}$ are such that $b^2 < 4c$. Then*

$$T^2 + bT + cI$$

*is an invertible operator.*

*Proof.* Let $v$ be a nonzero vector in $V$. Then

$$\begin{aligned}
\langle (T^2 + bT + cI)v, v\rangle &= \langle T^2 v, v\rangle + b\langle Tv, v\rangle + c\langle v, v\rangle \\
&= \langle Tv, Tv\rangle + b\langle Tv, v\rangle + c\|v\|^2 \\
&\geq \|Tv\|^2 - |b|\|Tv\|\|v\| + c\|v\|^2 \\
&= \left(\|Tv\| - \frac{|b|\|v\|}{2}\right)^2 + \left(c - \frac{b^2}{4}\right)\|v\|^2 \\
&> 0,
\end{aligned}$$

where the third line above holds by the Cauchy–Schwarz inequality. The last inequality implies that $(T^2 + bT + cI)v \neq 0$. Thus $T^2 + bT + cI$ is injective, which implies that it is invertible. $\dashv$

**Theorem 7.18.** *Suppose $T \in \mathcal{L}(V)$ is self-adjoint. Then the minimal polynomial of $T$ equals*

$$(z - \lambda_1) \cdots (z - \lambda_m)$$

*for some $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$.*

*Proof.* First suppose $\mathbb{F} = \mathbb{C}$. The zeros of the minimal polynomial of $T$ are the eigenvalues of $T$ (by 5.17). All eigenvalues of $T$ are real (by 7.8). Thus 4.8 tells us that the minimal polynomial of $T$ has the desired form.

Now suppose $\mathbb{F} = \mathbb{R}$. By 4.10 there exist

$$\lambda_1, \dots, \lambda_m \in \mathbb{R} \quad \text{and} \quad b_1, \dots, b_N,\ c_1, \dots, c_N \in \mathbb{R}$$

with $b_k^2 < 4c_k$ for each $k$ such that the minimal polynomial of $T$ equals

$$(z - \lambda_1) \cdots (z - \lambda_m)(z^2 + b_1 z + c_1) \cdots (z^2 + b_N z + c_N), \tag{$*$}$$

where either $m$ or $N$ might equal $0$. Now

$$(T - \lambda_1 I) \cdots (T - \lambda_m I)(T^2 + b_1 T + c_1 I) \cdots (T^2 + b_N T + c_N I) = 0.$$

If $N > 0$, then we could multiply both sides of the equation above on the right by the inverse of $T^2 + b_N T + c_N I$ (by 7.17) to obtain a polynomial expression of $T$ that equals $0$, violating the minimality of the degree of the polynomial with this property. Thus we must have $N = 0$, as desired. ⊣

**Theorem 7.19.** *Suppose $\mathbb{F} = \mathbb{R}$ and $T \in \mathcal{L}(V)$. Then the following are equivalent.*

*(a) $T$ is self-adjoint.*

*(b) $T$ has a diagonal matrix with respect to some orthonormal basis of $V$.*

*(c) $V$ has an orthonormal basis consisting of eigenvectors of $T$.*

*Proof.* First suppose (a) holds, so $T$ is self-adjoint. 6.24 and 7.18 imply that $T$ has an upper-triangular matrix with respect to some orthonormal basis of $V$. With respect to this orthonormal basis, the matrix of $T^*$ is the transpose of the matrix of $T$. However, $T^* = T$. Thus the transpose of the matrix of $T$ equals the matrix of $T$. Hence the matrix of $T$ is a diagonal matrix with respect to the orthonormal basis. Thus (a) implies (b).

Conversely, now suppose (b) holds, so $T$ has a diagonal matrix with respect to some orthonormal basis of $V$. That diagonal matrix equals its transpose. Thus with respect to that basis, the matrix of $T^*$ equals the matrix of $T$. Hence $T^* = T$, proving that (b) implies (a).

The equivalence of (b) and (c) follows from the definitions. ⊣

**Theorem 7.20.** *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Then the following are equivalent.*

*(a) $T$ is normal.*

*(b) $T$ has a diagonal matrix with respect to some orthonormal basis of $V$.*

*(c) $V$ has an orthonormal basis consisting of eigenvectors of $T$.*

*Proof.* First suppose (a) holds, so $T$ is normal. By 6.25 there is an orthonormal basis $e_1, \ldots, e_n$ of $V$ with respect to which $T$ has an upper-triangular matrix. Thus we can write

$$\mathcal{M}(T, (e_1, \ldots, e_n)) = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ & \ddots & \vdots \\ 0 & & a_{n,n} \end{pmatrix}. \tag{$*$}$$

We see from the matrix above that

$$\|Te_1\|^2 = |a_{1,1}|^2, \qquad \|T^*e_1\|^2 = |a_{1,1}|^2 + |a_{1,2}|^2 + \cdots + |a_{1,n}|^2.$$

Because $T$ is normal, $\|Te_1\| = \|T^*e_1\|$. Thus the two equations above imply that all entries in the first row of the matrix in $(*)$, except possibly the first entry $a_{1,1}$, equal 0.

Now $(*)$ implies

$$\|Te_2\|^2 = |a_{2,2}|^2$$

(because $a_{1,2} = 0$, as we showed in the paragraph above) and

$$\|T^*e_2\|^2 = |a_{2,2}|^2 + |a_{2,3}|^2 + \cdots + |a_{2,n}|^2.$$

Because $T$ is normal, $\|Te_2\| = \|T^*e_2\|$. Thus the two equations above imply that all entries in the second row of the matrix in $(*)$, except possibly the diagonal entry $a_{2,2}$, equal 0.

Continuing in this fashion, we see that all nondiagonal entries in the matrix $(*)$ equal 0. Thus (b) holds, completing the proof that (a) implies (b).

Now suppose (b) holds, so $T$ has a diagonal matrix with respect to some orthonormal basis of $V$. The matrix of $T^*$ (with respect to the same basis) is obtained by taking the conjugate transpose of the matrix of $T$; hence $T^*$ also has a diagonal matrix. Any two diagonal matrices commute; thus $T$ commutes with $T^*$, which means that $T$ is normal. In other words, (a) holds, completing the proof that (b) implies (a).

The equivalence of (b) and (c) follows from the definitions. ⊣

## 7C Positive Operators

**Definition 7.21.** An operator $T \in \mathcal{L}(V)$ is called *positive* if $T$ is self-adjoint and

$$\langle Tv, v \rangle \geq 0$$

for all $v \in V$.

If $V$ is a complex vector space, then the requirement that $T$ be self-adjoint can be dropped from the definition above (by 7.10).

**Definition 7.22.** An operator $R$ is called a square root of an operator $T$ if $R^2 = T$.

The characterizations of the positive operators in the next result correspond to characterizations of the nonnegative numbers among $\mathbb{C}$.

**Theorem 7.23.** *Let $T \in \mathcal{L}(V)$. Then the following are equivalent.*

*(a) $T$ is a positive operator.*
*(b) $T$ is self-adjoint and all eigenvalues of $T$ are nonnegative.*
*(c) With respect to some orthonormal basis of $V$, the matrix of $T$ is a diagonal matrix with only nonnegative numbers on the diagonal.*
*(d) $T$ has a positive square root.*
*(e) $T$ has a self-adjoint square root.*
*(f) $T = R^*R$ for some $R \in \mathcal{L}(V)$.*

**Theorem 7.24.** *Every positive operator on $V$ has a unique positive square root.*

*Proof.* Suppose $T \in \mathcal{L}(V)$ is positive. Suppose $v \in V$ is an eigenvector of $T$. Hence there exists a real number $\lambda \geq 0$ such that $Tv = \lambda v$.

Let $R$ be a positive square root of $T$. Note that the spectral theorem asserts that there is an orthonormal basis $e_1, \dots, e_n$ of $V$ consisting of eigenvectors of $R$. Because $R$ is a positive operator, all its eigenvalues are nonnegative. Thus there exist nonnegative numbers $\lambda_1, \dots, \lambda_n$ such that

$$Re_k = \sqrt{\lambda_k}\, e_k \quad \text{for each } k = 1, \dots, n.$$

Because $e_1, \dots, e_n$ is a basis of $V$, we can write

$$v = a_1 e_1 + \cdots + a_n e_n$$

for some numbers $a_1, \dots, a_n \in \mathbb{F}$. Thus

$$\lambda v = Tv = R^2 v = a_1 \lambda_1 e_1 + \cdots + a_n \lambda_n e_n.$$

The equation above implies that

$$a_1\lambda e_1 + \cdots + a_n\lambda e_n = a_1\lambda_1 e_1 + \cdots + a_n\lambda_n e_n.$$

Thus $a_k(\lambda - \lambda_k) = 0$ for each $k = 1, \dots, n$. Hence

$$v = \sum_{\{k : \lambda_k = \lambda\}} a_k e_k.$$

Thus

$$Rv = \sum_{\{k : \lambda_k = \lambda\}} a_k \sqrt{\lambda_k} e_k = \sqrt{\lambda} v.$$

Hence the behavior of $R$ on the eigenvectors of $T$ is uniquely determined. By spectral theorem there is a basis of $V$ consisting of eigenvectors of $T$, done. $\dashv$

For $T$ a positive operator, $\sqrt{T}$ denotes the unique positive square root of $T$.

**Theorem 7.25.** *Suppose $T$ is a positive operator on $V$ and $v \in V$ is such that $\langle Tv, v\rangle = 0$. Then $Tv = 0$.*

## 7D Isometries, Unitary Operators, and Matrix Factorization

**Definition 7.26.** A linear map $S \in \mathcal{L}(V, W)$ is called an *isometry* if

$$\|Sv\| = \|v\|$$

for every $v \in V$.

Hence isometry preserves norms, and is injective (it maps only 0 to 0).

**Theorem 7.27.** *Suppose $S \in \mathcal{L}(V, W)$. Suppose $e_1, \dots, e_n$ is an orthonormal basis of $V$ and $f_1, \dots, f_m$ is an orthonormal basis of $W$. Then the following are equivalent.*

*(a) $S$ is an isometry.*

*(b) $S^*S = I$.*

*(c) $\langle Su, Sv\rangle = \langle u, v\rangle$ for all $u, v \in V$.*

*(d) $Se_1, \dots, Se_n$ is an orthonormal list in $W$.*

*(e) The columns of $\mathcal{M}\big(S, (e_1, \dots, e_n), (f_1, \dots, f_m)\big)$ form an orthonormal list in $\mathbb{F}^m$ with respect*

*to the Euclidean inner product.*

*Proof.* $(a) \Rightarrow (b)$ uses 7.11. We then prove only that $(e) \Rightarrow (a)$.

To show that $S$ is an isometry, suppose $v \in V$. Then

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n \tag{$*$}$$

and

$$\|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2.$$

Applying $S$ to both sides of $(*)$ gives

$$Sv = \langle v, e_1 \rangle Se_1 + \cdots + \langle v, e_n \rangle Se_n.$$

Thus

$$\|Sv\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, , e_n \rangle|^2 = \|v\|.$$ ⊣

**Definition 7.28.** An operator $S \in \mathcal{L}(V)$ is called *unitary* if $S$ is an invertible isometry.

The word "invertible" could be deleted, for that every isometry on a finite dimensional vector space is invertible.

**Theorem 7.29.** *Suppose $S \in \mathcal{L}(V)$. Suppose $e_1, \ldots, e_n$ is an orthonormal basis of $V$. Then the following are equivalent.*

(a) *$S$ is a unitary operator.*
(b) *$S^*S = SS^* = I$.*
(c) *$S$ is invertible and $S^{-1} = S^*$.*
(d) *$Se_1, \ldots, Se_n$ is an orthonormal basis of $V$.*
(e) *The rows of $\mathcal{M}(S, (e_1, \ldots, e_n))$ form an orthonormal basis of $\mathbb{F}^n$ with respect to the Euclidean inner product.*
(f) *$S^*$ is a unitary operator.*

*Proof.* We prove only $(a) \Rightarrow (b)$: By 7.27 $S^*S = I$. Multiply both sides of this equation by $S^{-1}$ to get $S^* = S^{-1}$. Hence $SS^* = SS^{-1} = I$. ⊣

Consider 7.7, and note that unit circle in $\mathbb{C}$ corresponds to the set of unitary operators.

**Theorem 7.30.** *Suppose $\lambda$ is an eigenvalue of a unitary operator. Then $|\lambda| = 1$.*

**Theorem 7.31.** *Suppose $\mathbb{F} = \mathbb{C}$ and $S \in \mathcal{L}(V)$. Then the following are equivalent.*

*(a) $S$ is a unitary operator.*

*(b) There is an orthonormal basis of $V$ consisting of eigenvectors of $S$ whose corresponding eigenvalues all have absolute value* 1.

This is trivial, by 7.20.

When starting with $n$-by-$n$ matrices instead of operators, unless otherwise specified assume that the associated operators live on $\mathbb{F}^n$ (with the Euclidean inner product) and that their matrices are computed with respect to the standard basis of $\mathbb{F}^n$.

**Definition 7.32.** An $n$-by-$n$ matrix is called unitary if its columns form an orthonormal list in $\mathbb{F}^n$.

**Theorem 7.33.** ***QR factorization*** *Suppose $A$ is a square matrix with linearly independent columns. Then there exist unique matrices $Q$ and $R$ such that $Q$ is unitary, $R$ is upper triangular with only positive numbers on its diagonal, and*

$$A = QR.$$

*Proof.* Let $v_1, \dots, v_n$ denote the columns of $A$, thought of as elements of $\mathbb{F}^n$. Apply the Gram–Schmidt procedure 6.21 to the list $v_1, \dots, v_n$, getting an orthonormal basis $e_1, \dots, e_n$ of $\mathbb{F}^n$ such that

$$\operatorname{span}(v_1, \dots, v_k) = \operatorname{span}(e_1, \dots, e_k) \tag{$*$}$$

for each $k = 1, \dots, n$. Let $R$ be the $n$-by-$n$ matrix defined by

$$R_{j,k} = \langle v_k, e_j \rangle,$$

where $R_{j,k}$ denotes the entry in row $j$, column $k$ of $R$. If $j > k$, then $e_j$ is orthogonal to $\operatorname{span}(e_1, \dots, e_k)$ and hence $e_j$ is orthogonal to $v_k$ by ($*$). In other words, if $j > k$ then $\langle v_k, e_j \rangle = 0$. Thus $R$ is an upper-triangular matrix.

Let $Q$ be the unitary matrix whose columns are $e_1, \dots, e_n$. If $k \in \{1, \dots, n\}$, then the $k$th column of $QR$ equals a linear combination of the columns of $Q$, with the coefficients for the linear combination coming from the $k$th column of $R$. Hence the $k$th column of $QR$ equals

$$\langle v_k, e_1 \rangle e_1 + \cdots + \langle v_k, e_k \rangle e_k,$$

which equals $v_k$, the $k$th column of $A$. Thus $A = QR$, as desired.

The equations defining the Gram–Schmidt procedure show that each $v_k$ equals a positive multiple of $e_k$ plus a linear combination of $e_1, \dots, e_{k-1}$. Thus each $\langle v_k, e_k \rangle$ is a positive number. Hence all entries on the diagonal of $R$ are positive numbers.

Suppose we have $A = \widehat{Q}\widehat{R}$, where $\widehat{Q}$ is unitary and $\widehat{R}$ is upper triangular with only positive numbers on its diagonal. Let $q_1, \dots, q_n$ denote the columns of $\widehat{Q}$. Note that each $v_k$ is a linear combination of $q_1, \dots, q_k$, with the coefficients coming from the $k$th column of $\widehat{R}$. This implies that

$$\operatorname{span}(v_1, \dots, v_k) = \operatorname{span}(q_1, \dots, q_k)$$

and $\langle v_k, q_k \rangle > 0$. By 6B.10 $q_k = e_k$ for each $k = 1, \dots, n$. Hence $\widehat{Q} = Q$, which then implies that $\widehat{R} = R$. ⊣

Suppose $A$ is an $n$-by-$n$ square matrix with linearly independent columns. Suppose that $b \in \mathbb{F}^n$ and we want to solve the equation $Ax = b$ for $x = (x_1, \dots, x_n) \in \mathbb{F}^n$. Suppose $A = QR$, where $Q$ and $R$ is computed as in 7.33, then $Rx = Q^* b$, which can be quickly solved.

**Theorem 7.34.** *A self-adjoint operator $T \in \mathcal{L}(V)$ is a positive invertible operator if and only if $\langle Tv, v \rangle > 0$ for every nonzero $v \in V$.*

The proof uses 7.25.

**Definition 7.35.** A matrix $B \in \mathbb{F}^{n,n}$ is called *positive definite* if $B^* = B$ and

$$\langle Bx, x \rangle > 0$$

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

**Theorem 7.36.** ***Cholesky factorization*** *Suppose $B$ is a positive definite matrix. Then there exists a unique upper-triangular matrix $R$ with only positive numbers on its diagonal such that*

$$B = R^* R.$$

*Proof.* Because $B$ is positive definite, there exists an invertible square matrix $A$ of the same size as $B$ such that

$$B = A^* A$$

Let $A = QR$ be the QR factorization of $A$, where $Q$ is unitary and $R$ is upper triangular with only positive numbers on its diagonal. Then $A^* = R^*Q^*$. Thus

$$B = A^*A = R^*Q^*QR = R^*R.$$

To prove the uniqueness part of this result, suppose $S$ is an upper-triangular matrix with only positive numbers on its diagonal and $B = S^*S$. The matrix $S$ is invertible because $B$ is invertible (3D.11). Multiplying both sides of the equation $B = S^*S$ by $S^{-1}$ on the left gives the equation

$$BS^{-1} = S^*.$$

Let $A$ be the matrix from the first paragraph of this proof. Then

$$\begin{aligned}(AS^{-1})^*(AS^{-1}) &= (S^{-1})^*A^*AS^{-1}\\ &= (S^*)^{-1}BS^{-1}\\ &= (S^*)^{-1}S^*\\ &= I.\end{aligned}$$

Thus $AS^{-1}$ is unitary. Hence

$$A = (AS^{-1})S$$

is a factorization of $A$ as the product of a unitary matrix and an upper-triangular matrix with only positive numbers on its diagonal. The uniqueness of the QR factorization implies that $S = R$. ⊣

## 7E Singular Value Decomposition

**Theorem 7.37.** *Suppose $T \in \mathcal{L}(V, W)$. Then*

*(a)* $T^*T$ *is a positive operator on* $V$*;*

*(b)* $\operatorname{null} T^*T = \operatorname{null} T$*;*

*(c)* $\operatorname{range} T^*T = \operatorname{range} T^*$*;*

*(d)* $\dim \operatorname{range} T = \dim \operatorname{range} T^* = \dim \operatorname{range} T^*T$.

*Proof.* We prove (c) only. By 7.4,

$$\operatorname{range} T^*T = \left(\operatorname{null} T^*T\right)^{\perp} = \left(\operatorname{null} T\right)^{\perp} = \operatorname{range} T^*.$$ ⊣

**Definition 7.38.** Suppose $T \in \mathcal{L}(V, W)$. The *singular values* of $T$ are the nonnegative square roots of the eigenvalues of $T^*T$, listed in decreasing order, each included as many times as the dimension of the corresponding eigenspace of $T^*T$.

**Theorem 7.39.** *Suppose that $T \in \mathcal{L}(V, W)$. Then*

*(a) $T$ is injective $\iff$ $0$ is not a singular value of $T$;*

*(b) the number of positive singular values of $T$ equals $\dim \operatorname{range} T$;*

*(c) $T$ is surjective $\iff$ the number of positive singular values of $T$ equals $\dim W$.*

These are trivial.

**Theorem 7.40.** *Suppose that $S \in \mathcal{L}(V, W)$. Then*

$$S \text{ is an isometry} \iff \text{all singular values of } S \text{ equal } 1.$$

**Theorem 7.41.** ***SVD*** *Suppose $T \in \mathcal{L}(V, W)$ and the positive singular values of $T$ are $s_1, \dots, s_m$. Then there exist orthonormal lists $e_1, \dots, e_m$ in $V$ and $f_1, \dots, f_m$ in $W$ such that*

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m$$

*for every $v \in V$.*

*Proof.* Let $s_1, \dots, s_n$ denote the singular values of $T$ (thus $n = \dim V$). Because $T^*T$ is a positive operator, the spectral theorem implies that there exists an orthonormal basis $e_1, \dots, e_n$ of $V$ with

$$T^*Te_k = s_k^2 e_k$$

for each $k = 1, \dots, n$. For each $k = 1, \dots, m$, let

$$f_k = \frac{Te_k}{s_k}.$$

If $j, k \in \{1, \dots, m\}$, then

$$\langle f_j, f_k\rangle = \frac{1}{s_j s_k}\langle Te_j, Te_k\rangle = \frac{1}{s_j s_k}\langle e_j, T^*Te_k\rangle = \frac{s_k}{s_j}\langle e_j, e_k\rangle = \begin{cases} 0 & \text{if } j \neq k, \\ 1 & \text{if } j = k. \end{cases}$$

Thus $f_1, \ldots, f_m$ is an orthonormal list in $W$. If $k \in \{1, \ldots, n\}$ and $k > m$, then $s_k = 0$ and hence $T^*Te_k = 0$, thus $Te_k = 0$ by 7.37. Suppose $v \in V$. Then

$$\begin{aligned} Tv &= T\big(\langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n\big) \\ &= \langle v, e_1\rangle Te_1 + \cdots + \langle v, e_m\rangle Te_m \\ &= s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m. \end{aligned}$$

⊣

**Definition 7.42.** An $M$-by-$N$ matrix $A$ is called a *diagonal matrix* if all entries of the matrix are 0 except possibly $A_{k,k}$ for $k = 1, \ldots, \min\{M, N\}$.

**Remark 7.43.** Suppose $T \in \mathcal{L}(V, W)$, the positive singular values of $T$ are $s_1, \ldots, s_m$, and $e_1, \ldots, e_m$ and $f_1, \ldots, f_m$ are as in 7.41. The orthonormal list $e_1, \ldots, e_m$ can be extended to an orthonormal basis $e_1, \ldots, e_{\dim V}$ of $V$ and the orthonormal list $f_1, \ldots, f_m$ can be extended to an orthonormal basis $f_1, \ldots, f_{\dim W}$ of $W$. Note that

$$Te_k = \begin{cases} s_k f_k & \text{if } 1 \leq k \leq m, \\ 0 & \text{if } m < k \leq \dim V. \end{cases}$$

Thus $T$ has a diagonal matrix

$$\mathcal{M}\big(T, (e_1, \ldots, e_{\dim V}), (f_1, \ldots, f_{\dim W})\big)_{j,k} = \begin{cases} s_k & \text{if } 1 \leq j = k \leq m, \\ 0 & \text{otherwise.} \end{cases}$$

**Theorem 7.44.** *Suppose $T \in \mathcal{L}(V, W)$ and the positive singular values of $T$ are $s_1, \ldots, s_m$. Suppose $e_1, \ldots, e_m$ and $f_1, \ldots, f_m$ are orthonormal lists in $V$ and $W$ such that*

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m \tag{$*$}$$

*for every $v \in V$. Then*

$$T^*w = s_1\langle w, f_1\rangle e_1 + \cdots + s_m\langle w, f_m\rangle e_m$$

*and*

$$T^\dagger w = \frac{\langle w, f_1\rangle}{s_1} e_1 + \cdots + \frac{\langle w, f_m\rangle}{s_m} e_m$$

*for every $w \in W$.*

*Proof.* If $v \in V$ and $w \in W$ then

$$\begin{aligned}\langle Tv, w\rangle &= \langle s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m, w\rangle \\ &= s_1\langle v, e_1\rangle\langle f_1, w\rangle + \cdots + s_m\langle v, e_m\rangle\langle f_m, w\rangle \\ &= \langle v, s_1\langle w, f_1\rangle e_1 + \cdots + s_m\langle w, f_m\rangle e_m\rangle,\end{aligned}$$

giving the adjoint of $T$. For the pseudo inverse (6.38), suppose $w \in W$. Let

$$v = \frac{\langle w, f_1\rangle}{s_1} e_1 + \cdots + \frac{\langle w, f_m\rangle}{s_m} e_m.$$

Apply $T$ to both sides of the equation above, getting

$$\begin{aligned}Tv &= \frac{\langle w, f_1\rangle}{s_1} Te_1 + \cdots + \frac{\langle w, f_m\rangle}{s_m} Te_m \\ &= \langle w, f_1\rangle f_1 + \cdots + \langle w, f_m\rangle f_m \\ &= P_{\mathrm{range}\,T} w,\end{aligned}$$

where the second line holds because $(*)$ implies that $Te_k = s_k f_k$ if $k = 1, \ldots, m$, and the last line holds because $(*)$ implies that $f_1, \ldots, f_m$ spans range $T$ and thus is an orthonormal basis of range $T$. By 7E.8 (b) $v \in (\mathrm{null}\,T)^\perp$. Thus $v = T^\dagger w$. ⊣

**Theorem 7.45.** ***SVD, matrix version*** *Suppose $A$ is an $M$-by-$n$ matrix of rank $m \geq 1$. Then there exist an $M$-by-$m$ matrix $B$ with orthonormal columns, an $m$-by-$m$ diagonal matrix $D$ with positive numbers on the diagonal, and an $n$-by-$m$ matrix $C$ with orthonormal columns such that*

$$A = BDC^*.$$

*Proof.* Let $T : \mathbb{F}^n \to \mathbb{F}^M$ be the linear map whose matrix with respect to the standard bases equals $A$. Then $\dim \mathrm{range}\,T = m$. Let

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m \tag{$*$}$$

be an SVD of $T$. Let

$$\begin{aligned}B &= \text{the } M\text{-by-}m \text{ matrix whose columns are } f_1, \ldots, f_m, \\ C &= \text{the } m\text{-by-}m \text{ diagonal matrix whose diagonal entries are } s_1, \ldots, s_m, \\ D &= \text{the } n\text{-by-}m \text{ matrix whose columns are } e_1, \ldots, e_m.\end{aligned}$$

Let $u_1, \ldots, u_m$ denote the standard basis of $\mathbb{F}^m$. If $k \in \{1, \ldots, m\}$ then

$$(AC - BD)u_k = Ae_k - B(s_k u_k) = s_k f_k - s_k f_k = 0.$$

Thus $AC = BD$. Multiply both sides of this last equation by $C^*$ (the conjugate transpose of $C$) on the right to get

$$ACC^* = BDC^*.$$

If $k \in \{1, \dots, m\}$, then $C^* e_k = u_k$; hence $CC^* e_k = e_k$. Thus $ACC^* v = Av$ for all $v \in \operatorname{span}(e_1, \dots, e_m)$. If $v \in \operatorname{span}(e_1, \dots, e_m)^\perp$, then $Av = 0$ (as follows from $(*)$) and $C^* v = 0$. Hence $ACC^* v = Av$ for all $v \in \operatorname{span}(e_1, \dots, e_m)^\perp$. We conclude that $ACC^* = A$, as desired. ⊣

*Comment.* The matrix $A$ has $Mn$ entries. In comparison, the matrices $B$, $D$, and $C$ above have a total of

$$m(M + m + n)$$

entries, probably considerably less then $Mn$. Also, by 7.44 we can easily have

$$A^\dagger = CD^{-1}B^*.$$

## Exercises

**Exercise 8.** Suppose $T \in \mathcal{L}(V, W)$. Suppose $s_1 \geq s_2 \geq \cdots \geq s_m > 0$ and $e_1, \dots, e_m$ is an orthonormal list in $V$ and $f_1, \dots, f_m$ is an orthonormal list in $W$ such that

$$Tv = s_1 \langle v, e_1 \rangle f_1 + \cdots + s_m \langle v, e_m \rangle f_m$$

for every $v \in V$.

(a) Prove that $f_1, \dots, f_m$ is an orthonormal basis of range $T$.
(b) Prove that $e_1, \dots, e_m$ is an orthonormal basis of $(\operatorname{null} T)^\perp$.
(c) Prove that $s_1, \dots, s_m$ are the positive singular values of $T$.
(d) Prove that if $k \in \{1, \dots, m\}$, then $e_k$ is an eigenvector of $T^*T$ with corresponding eigenvalue $s_k^2$.
(e) Prove that

$$TT^* w = s_1^2 \langle w, f_1 \rangle f_1 + \cdots + s_m^2 \langle w, f_m \rangle f_m$$

for all $w \in W$.

(a) and (b) are trivial. (c), (d), and (e) are by $T^* w$ in 7.44.

**Exercise 9.** Suppose $T \in \mathcal{L}(V, W)$. Show that $T$ and $T^*$ have the same positive singular values.

7E.8 (c) we have that positive singular values of $T$ are uniquely determined given an SVD of $T$. By 7.44 we have that $T^*$ and $T$ share positive singular values.

**Exercise 10.** Suppose $T \in \mathcal{L}(V, W)$ has singular values $s_1, \dots, s_n$. Prove that if $T$ is an invertible linear map, then $T^{-1}$ has singular values

$$\frac{1}{s_n}, \dots, \frac{1}{s_1}.$$

If $T$ is invertible, we have that $T^{-1} = T^{\dagger}$, by 7E.8 and 7.44 we are done.

## 7F Consequences of SVD

**Theorem 7.46.** *Suppose $T \in \mathcal{L}(V, W)$. Let $s_1$ be the largest singular value of $T$. Then*

$$\|Tv\| \leq s_1 \|v\|$$

*for all $v \in V$.*

*Proof.* By 7.41,

$$\|Tv\|^2 = \sum_{i=1}^{m} s_i^2 \langle v, e_i \rangle^2 \leq s_1^2 \sum_{i=1}^{m} \langle v, e_i \rangle^2 \leq s_1^2 \|v\|^2. \qquad \dashv$$

**Definition 7.47.** Suppose $T \in \mathcal{L}(V, W)$. Then the *norm* of $T$, denoted by $\|T\|$, is defined by

$$\|T\| = \max\{\|Tv\| \,:\, v \in V \text{ and } \|v\| \leq 1\}.$$

**Theorem 7.48.** *Suppose $T \in \mathcal{L}(V, W)$. Then*

*(a)* $\|T\| \geq 0$;

*(b)* $\|T\| = 0 \iff T = 0$;

*(c)* $\|\lambda T\| = |\lambda| \|T\|$ *for all* $\lambda \in \mathbb{F}$;

*(d)* $\|S + T\| \leq \|S\| + \|T\|$ *for all* $S \in \mathcal{L}(V, W)$.

**Theorem 7.49.** *Suppose $T \in \mathcal{L}(V, W)$. Then*

*(a)* $\|T\|$ = *the largest singular value of* $T$;

*(b)* $\|T\| = \max\{\|Tv\| \,:\, v \in V$ *and* $\|v\| = 1\}$;

*(c)* $\|T\| =$ *the smallest number* $c$ *such that* $\|Tv\| \leq c\|v\|$ *for all* $v \in V$.

**Theorem 7.50.** *Suppose* $T \in \mathcal{L}(V, W)$*. Then* $\|T\| = \|T^*\|$.

This is immediate by 7E.9.

Recall 7.7 and note that every $z \in \mathbb{C}$ can be written in the form

$$z = \frac{z}{|z|}\sqrt{\bar{z}z}.$$

The following theorem is very similar.

**Theorem 7.51.** ***Polar Decomposition*** *Suppose* $T \in \mathcal{L}(V)$*. Then there exists a unitary operator* $S \in \mathcal{L}(V)$ *such that*

$$T = S\sqrt{T^*T}.$$

*Proof.* Let $s_1, \dots, s_m$ be the positive singular values of $T$, and let $e_1, \dots, e_m$ and $f_1, \dots, f_m$ be orthonormal lists in $V$ such that

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m \quad (v \in V).$$

Extend $e_1, \dots, e_m$ and $f_1, \dots, f_m$ to orthonormal bases $e_1, \dots, e_n$ and $f_1, \dots, f_n$ of $V$. Define $S \in \mathcal{L}(V)$ by

$$Sv = \langle v, e_1\rangle f_1 + \cdots + \langle v, e_n\rangle f_n \quad (v \in V).$$

Then

$$\|Sv\|^2 = \|\langle v, e_1\rangle f_1 + \cdots + \langle v, e_n\rangle f_n\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2 = \|v\|^2,$$

hence $S$ is unitary. Applying $T^*$ to both sides of the first display and using 7.44 for $T^*$ gives

$$T^*Tv = s_1^2\langle v, e_1\rangle e_1 + \cdots + s_m^2\langle v, e_m\rangle e_m \quad (v \in V).$$

Thus the positive operator $R$ defined by

$$Rv = s_1\langle v, e_1\rangle e_1 + \cdots + s_m\langle v, e_m\rangle e_m$$

satisfies $R^2 = T^*T$, hence $R = \sqrt{T^*T}$. Now

$$\begin{aligned} S\sqrt{T^*T}\,v &= S(Rv) \\ &= S\big(s_1\langle v, e_1\rangle e_1 + \cdots + s_m\langle v, e_m\rangle e_m\big) \\ &= s_1\langle v, e_1\rangle f_1 + \cdots + s_m\langle v, e_m\rangle f_m \\ &= Tv, \end{aligned}$$

where the last equality is the first display. Therefore $T = S\sqrt{T^*T}$. ⊣

Chapter 8

# Operators on Complex Vector Spaces

## 8A Generalized Eigenvectors and Nilpotent Operators

**Theorem 8.1.** *Suppose $T \in \mathcal{L}(V)$. Then*

$$\{0\} = \operatorname{null} T^0 \subseteq \operatorname{null} T^1 \subseteq \cdots \subseteq \operatorname{null} T^k \subseteq \operatorname{null} T^{k+1} \subseteq \cdots.$$

**Theorem 8.2.** *Suppose $T \in \mathcal{L}(V)$ and $m$ is a nonnegative integer such that*

$$\operatorname{null} T^m = \operatorname{null} T^{m+1}.$$

*Then*

$$\operatorname{null} T^m = \operatorname{null} T^{m+1} = \operatorname{null} T^{m+2} = \operatorname{null} T^{m+3} = \cdots.$$

**Theorem 8.3.** *Suppose $T \in \mathcal{L}(V)$. Then*

$$\operatorname{null} T^{\dim V} = \operatorname{null} T^{\dim V+1} = \operatorname{null} T^{\dim V+2} = \cdots.$$

*Proof.* We only need to prove that $\operatorname{null} T^{\dim V} = \operatorname{null} T^{\dim V+1}$ by 8.2. Suppose this is not true. Then, by 8.1 and 8.2, we have

$$\{0\} = \operatorname{null} T^0 \subsetneq \operatorname{null} T^1 \subsetneq \cdots \subsetneq \operatorname{null} T^{\dim V} \subsetneq \operatorname{null} T^{\dim V+1},$$

At each of the strict inclusions in the chain above, the dimension increases by at least 1. Thus $\dim \operatorname{null} T^{\dim V+1} \geq \dim V + 1$, a contradiction because a subspace of $V$ cannot have a larger dimension than $\dim V$. ⊣

**Theorem 8.4.** *Suppose $T \in \mathcal{L}(V)$. Then*

$$V = \operatorname{null} T^{\dim V} \oplus \operatorname{range} T^{\dim V}.$$

The proof is by 3.15 and 1.18.

Some operators do not have enough eigenvectors to lead to good descriptions of their behavior.

**Definition 8.5.** Suppose $T \in \mathcal{L}(V)$ and $\lambda$ is an eigenvalue of $T$. A vector $v \in V$ is called a *generalized eigenvector* of $T$ corresponding to $\lambda$ if $v \neq 0$ and

$$(T - \lambda I)^k v = 0$$

for some positive integer $k$.

**Theorem 8.6.** *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Then there is a basis of $V$ consisting of generalized eigenvectors of $T$.*

*Proof.* Let $n = \dim V$. Use induction on $n$. To start, note that the desired result holds if $n = 1$, because then every nonzero vector in $V$ is an eigenvector of $T$.

Now suppose $n > 1$ and the desired result holds for all smaller values of $\dim V$. Let $\lambda$ be an eigenvalue of $T$. Applying 8.4 to $T - \lambda I$ shows that

$$V = \operatorname{null}((T - \lambda I)^n) \oplus \operatorname{range}((T - \lambda I)^n).$$

If $\operatorname{null}((T - \lambda I)^n) = V$, then every nonzero vector in $V$ is a generalized eigenvector of $T$, so in this case there is a basis of $V$ consisting of generalized eigenvectors of $T$. Hence we can assume that $\operatorname{null}((T - \lambda I)^n) \neq V$, which implies that $\operatorname{range}((T - \lambda I)^n) \neq \{0\}$. Also, $\operatorname{null}((T - \lambda I)^n) \neq \{0\}$ because $\lambda$ is an eigenvalue of $T$. Thus we have

$$0 < \dim \operatorname{range}((T - \lambda I)^n) < n.$$

Furthermore, $\operatorname{range}((T - \lambda I)^n)$ is invariant under $T$ (by 5.12). Let $S \in \mathcal{L}(\operatorname{range}((T - \lambda I)^n))$ be $T$ restricted to $\operatorname{range}((T - \lambda I)^n)$. Our induction hypothesis applied to $S$ implies that there is a basis of $\operatorname{range}((T - \lambda I)^n)$ consisting of generalized eigenvectors of $S$, which are of course generalized eigenvectors of $T$. Adjoining that basis of $\operatorname{range}((T - \lambda I)^n)$ to a basis of $\operatorname{null}((T - \lambda I)^n)$ gives a basis of $V$ consisting of generalized eigenvectors of $T$. ⊣

**Theorem 8.7.** *Suppose $T \in \mathcal{L}(V)$. Then each generalized eigenvector of $T$ corresponds to only one eigenvalue of $T$.*

*Proof.* Suppose $v \in V$ is a generalized eigenvector of $T$ corresponding to eigenvalues $\alpha$ and $\lambda$ of $T$. Let $m$ be the smallest positive integer such that $(T - \alpha I)^m v = 0$. Let $n = \dim V$. Then

$$\begin{aligned} 0 &= (T - \lambda I)^n v \\ &= ((T - \alpha I) + (\alpha - \lambda) I)^n v \\ &= \sum_{k=0}^{n} b_k (\alpha - \lambda)^{n-k} (T - \alpha I)^k v, \end{aligned}$$

where $b_0 = 1$ and the values of the other binomial coefficients $b_k$ do not matter. Apply the operator $(T - \alpha I)^{m-1}$ to both sides of the equation above, obtaining

$$0 = (\alpha - \lambda)^n (T - \alpha I)^{m-1} v.$$

Because $(T - \alpha I)^{m-1} v \neq 0$, the equation above implies that $\alpha = \lambda$, as desired. ⊣

**Theorem 8.8.** *Suppose that $T \in \mathcal{L}(V)$. Then every list of generalized eigenvectors of $T$ corresponding to distinct eigenvalues of $T$ is linearly independent. Cf. 5.7.*

*Proof.* Suppose the desired result is false. Then there exists a smallest positive integer $m$ such that there is a linearly dependent list $v_1, \ldots, v_m$ of generalized eigenvectors of $T$ corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_m$ of $T$ (note that $m \geq 2$ because a generalized eigenvector is nonzero by definition). Thus there exist $a_1, \ldots, a_m \in \mathbb{F}$, none of which are 0 (by minimality of $m$), such that

$$a_1 v_1 + \cdots + a_m v_m = 0.$$

Let $n = \dim V$. Apply $(T - \lambda_m I)^n$ to both sides to obtain

$$a_1 (T - \lambda_m I)^n v_1 + \cdots + a_{m-1} (T - \lambda_m I)^n v_{m-1} = 0.$$

Suppose $k \in \{1, \ldots, m-1\}$. Then

$$(T - \lambda_m I)^n v_k \neq 0,$$

because otherwise $v_k$ would be a generalized eigenvector corresponding to both $\lambda_k$ and $\lambda_m$, contradicting 8.7. However,

$$(T - \lambda_k I)^n ((T - \lambda_m I)^n v_k) = (T - \lambda_m I)^n ((T - \lambda_k I)^n v_k) = 0.$$

Hence $(T - \lambda_m I)^n v_k$ is a generalized eigenvector of $T$ corresponding to the eigenvalue $\lambda_k$. Thus $(T - \lambda_m I)^n v_1, \ldots, (T - \lambda_m I)^n v_{m-1}$ is a linearly dependent list of $m - 1$ generalized eigenvectors corresponding to distinct eigenvalues, contradicting the minimality of $m$, as desired. ⊣

**Definition 8.9.** An operator is called *nilpotent* if some power of it equals $0$.

**Theorem 8.10.** *Suppose $T \in \mathcal{L}(V)$ is nilpotent. Then $T^{\dim V} = 0$.*

**Theorem 8.11.** *Suppose $T \in \mathcal{L}(V)$.*

*(a) If $T$ is nilpotent, then $0$ is an eigenvalue of $T$ and $T$ has no other eigenvalues.*

*(b) If $\mathbb{F} = \mathbb{C}$ and $0$ is the only eigenvalue of $T$, then $T$ is nilpotent.*

(b) is trivial by 5.17.

**Theorem 8.12.** *Suppose $T \in \mathcal{L}(V)$. Then the following are equivalent:*

*(a) $T$ is nilpotent.*

*(b) The minimal polynomial of $T$ is $z^m$ for some positive integer $m$.*

*(c) There exists a basis of $V$ with respect to which the matrix of $T$ has the strictly upper-triangular form*

$$\begin{pmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{pmatrix}.$$

This is trivial, by 5.19, 5.17, 5.27, 5.26, and 5.25.

# Exercise

**Exercise 2.** Suppose $T \in \mathcal{L}(V)$, $m$ is a positive integer, $v \in V$, and $T^{m-1}v \neq 0$ but $T^m v = 0$. Prove that $v, Tv, T^2v, \ldots, T^{m-1}v$ is linearly independent.

*Proof.* Suppose $\sum_{j=0}^{m-1} c_j T^j u = 0$. Applying $T^{n-1}$ yields $c_0 T^{n-1}u = 0$, so $c_0 = 0$ since $T^{n-1}u \neq 0$. Then applying $T^{n-2}$ to $\sum_{j=1}^{n-1} c_j T^j u = 0$ gives $c_1 T^{n-1} u = 0$, hence $c_1 = 0$. Repeating this argument inductively, we obtain $c_j = 0$ for all $j$. ⊣

# 8B Generalized Eigenspace Decomposition

**Definition 8.13.** Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. The *generalized eigenspace* of $T$ corresponding to $\lambda$, denoted by $G(\lambda, T)$, is defined by

$$G(\lambda, T) = \{v \in V : (T - \lambda I)^k v = 0 \text{ for some positive integer } k\}.$$

Thus $G(\lambda, T)$ is the set of generalized eigenvectors of $T$ corresponding to $\lambda$, together with the zero vector.

**Theorem 8.14.** *Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. Then $G(\lambda, T) = \operatorname{null}\big((T - \lambda I)^{\dim V}\big)$.*

**Theorem 8.15.** *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of $T$. Then*

*(a) $G(\lambda_k, T)$ is invariant under $T$ for each $k = 1, \dots, m$;*
*(b) $(T - \lambda_k I)|_{G(\lambda_k, T)}$ is nilpotent for each $k = 1, \dots, m$;*
*(c) $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$.*

*Proof.* (a) is by 5.12. (b) is by 8.14. (c) is by 8.8, 1.17, and 8.6. ⊣

**Definition 8.16.** Suppose $T \in \mathcal{L}(V)$. The *multiplicity* of an eigenvalue $\lambda$ of $T$ is defined to be the dimension of the corresponding generalized eigenspace $G(\lambda, T)$.

**Theorem 8.17.** *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Then the sum of the multiplicities of all eigenvalues of $T$ equals* $\dim V$.

Multiplicity is sometimes called *algebraic multiplicity*. And $\dim E(\lambda, T)$ is sometimes called *geometric multiplicity*.

If $V$ is an inner product space, $T \in \mathcal{L}(V)$ is normal, and $\lambda$ is an eigenvalue of $T$, then the algebraic multiplicity of $\lambda$ equals the geometric multiplicity of $\lambda$, by 7A.27.

**Definition 8.18.** Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ denote the distinct eigenvalues of $T$, with multiplicities $d_1, \dots, d_m$. The polynomial

$$(z - \lambda_1)^{d_1} \cdots (z - \lambda_m)^{d_m}$$

is called the *characteristic polynomial* of $T$.

**Theorem 8.19.** *Suppose* $\mathbb{F} = \mathbb{C}$ *and* $T \in \mathcal{L}(V)$. *Then*

*(a) the characteristic polynomial of* $T$ *has degree* $\dim V$*;*

*(b) the zeros of the characteristic polynomial of* $T$ *are the eigenvalues of* $T$*.*

**Theorem 8.20.** ***Cayley-Hamilton*** *Suppose* $\mathbb{F} = \mathbb{C}$, $T \in \mathcal{L}(V)$, *and* $q$ *is the characteristic polynomial of* $T$. *Then* $q(T) = 0$.

*Proof.* Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of $T$, and let $d_k = \dim G(\lambda_k, T)$. For each $k \in \{1, \dots, m\}$, $(T - \lambda_k I)\big|_{G(\lambda_k, T)}$ is nilpotent. Thus,

$$(T - \lambda_k I)^{d_k}\big|_{G(\lambda_k, T)} = 0 \qquad \text{for each } k \in \{1, \dots, m\}.$$

By 8.15 every vector in $V$ is a sum of vectors in

$$G(\lambda_1, T),\ G(\lambda_2, T),\ \dots,\ G(\lambda_m, T).$$

Hence, to prove $q(T) = 0$, it suffices to show

$$q(T)\big|_{G(\lambda_k, T)} = 0 \qquad \text{for each } k.$$

Fix $k \in \{1, \dots, m\}$. We have

$$q(T) = (T - \lambda_1 I)^{d_1} \cdots (T - \lambda_m I)^{d_m}.$$

The operators on the right–hand side commute with one another, so we may move the factor $(T - \lambda_k I)^{d_k}$ to the last position. Since

$$(T - \lambda_k I)^{d_k}\big|_{G(\lambda_k, T)} = 0,$$

it follows immediately that

$$q(T)\big|_{G(\lambda_k, T)} = 0,$$

as desired. ⊣

**Corollary 8.21.** *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Then the characteristic polynomial of $T$ is a polynomial multiple of the minimal polynomial of $T$.*

This is by 5.19.

**Theorem 8.22.** *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Suppose $v_1, \ldots, v_n$ is a basis of $V$ such that $\mathcal{M}(T, (v_1, \ldots, v_n))$ is upper triangular. Then the number of times that each eigenvalue $\lambda$ of $T$ appears on the diagonal of $\mathcal{M}(T, (v_1, \ldots, v_n))$ equals the multiplicity of $\lambda$ as an eigenvalue of $T$.*

*Proof.* Let $A = \mathcal{M}(T, (v_1, \ldots, v_n))$ be an upper−triangular matrix. Let $\lambda_1, \ldots, \lambda_n$ denote the entries on the diagonal of $A$. Then for each $k \in \{1, \ldots, n\}$ we have

$$Tv_k = u_k + \lambda_k v_k$$

for some $u_k \in \operatorname{span}(v_1, \ldots, v_{k-1})$. Hence, if $k \in \{1, \ldots, n\}$ and $\lambda_k \neq 0$, then $Tv_k$ is not a linear combination of $Tv_1, \ldots, Tv_{k-1}$ and the list of those $Tv_k$ such that $\lambda_k \neq 0$ is linearly independent.

Let $d$ denote the number of indices $k \in \{1, \ldots, n\}$ such that $\lambda_k = 0$. Then

$$\dim \operatorname{range} T \geq n - d.$$

Because $n = \dim V = \dim \operatorname{null} T + \dim \operatorname{range} T$, the inequality above gives

$$\dim \operatorname{null} T \leq d. \tag{$*$}$$

By 5C.2 the matrix of the operator $T^n$ with respect to the basis $v_1, \ldots, v_n$ is the upper−triangular matrix $A^n$, whose diagonal entries are $\lambda_1^n, \ldots, \lambda_n^n$. Because $\lambda_k^n = 0$ iff $\lambda_k = 0$, the number of times 0 appears on the diagonal of $A^n$ equals $d$. Thus, applying $(*)$ with $T$ replaced by $T^n$, we obtain

$$\dim \operatorname{null} T^n \leq d. \tag{$**$}$$

Now, for an eigenvalue $\lambda$ of $T$, let $m_\lambda$ denote the multiplicity of $\lambda$ as an eigenvalue of $T$, and let $d_\lambda$ denote the number of times $\lambda$ appears on the diagonal of $A$. Replacing $T$ by $T - \lambda I$ in $(**)$ we have

$$m_\lambda \leq d_\lambda$$

for each eigenvalue $\lambda$ of $T$.

The sum of the multiplicities $m_\lambda$ over all eigenvalues $\lambda$ of $T$ equals $n = \dim V$. The sum of the $d_\lambda$ over all eigenvalues $\lambda$ of $T$ also equals $n$, since the diagonal of $A$ has length $n$. Hence above must in fact be an equality for each $\lambda$. Hence, the multiplicity of $\lambda$ as an eigenvalue of $T$ equals the number of times $\lambda$ appears on the diagonal of $A$, as desired. ⊣

**Definition 8.23.** A *block diagonal matrix* is a square matrix of the form

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix},$$

where $A_1, \dots, A_m$ are square matrices lying along the diagonal, and all other entries of the matrix are 0.

**Theorem 8.24.** *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of $T$, with multiplicities $d_1, \dots, d_m$. Then there exists a basis of $V$ with respect to which $T$ has a block diagonal matrix of the form*

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix},$$

*where each $A_k$ is a $d_k \times d_k$ upper–triangular matrix of the form*

$$A_k = \begin{pmatrix} \lambda_k & & * \\ & \ddots & \\ 0 & & \lambda_k \end{pmatrix}.$$

*Proof.* Each $(T - \lambda_k I)\big|_{G(\lambda_k, T)}$ is nilpotent. For each $k$, choose a basis of $G(\lambda_k, T)$, which is a vector space of dimension $d_k$, such that the matrix of $(T - \lambda_k I)\big|_{G(\lambda_k, T)}$ with respect to this basis is as in 8.12(c). Thus, with respect to this basis, the matrix of $T\big|_{G(\lambda_k, T)}$, which equals

$$(T - \lambda_k I)\big|_{G(\lambda_k, T)} + \lambda_k I\big|_{G(\lambda_k, T)},$$

looks like the desired form shown above for $A_k$.

By 8.15 putting together the bases of the $G(\lambda_k, T)$'s chosen above gives a basis of $V$. The matrix of $T$ with respect to this basis has the desired form. ⊣

## 8C Consequences of Generalized Eigenspace Decomposition

**Theorem 8.25.** *Suppose $T \in \mathcal{L}(V)$ is nilpotent. Then $I + T$ has a square root.*

*Proof.* Consider as motivation the Taylor series for the function $\sqrt{1+x}$:

$$\sqrt{1+x} = 1 + a_1 x + a_2 x^2 + \cdots.$$

Because $T$ is nilpotent, $T^m = 0$ for some positive integer $m$. Thus we guess that there is a square root of $I + T$ of the form

$$I + a_1 T + a_2 T^2 + \cdots + a_{m-1} T^{m-1}.$$

Having made this guess, we can try to choose $a_1, a_2, \ldots, a_{m-1}$ such that the operator above has its square equal to $I + T$. Now,

$$\begin{aligned}(I+&a_1 T + a_2 T^2 + a_3 T^3 + \cdots + a_{m-1} T^{m-1})^2 \\ &= I + 2a_1 T + (2a_2 + a_1^2) T^2 + (2a_3 + 2a_1 a_2) T^3 + \cdots \\ &\quad + (2a_{m-1} + \text{terms involving } a_1, \ldots, a_{m-2}) T^{m-1}.\end{aligned}$$

We want the right side of the equation above to equal $I + T$. Hence, choose $a_1$ such that $2a_1 = 1$ (thus $a_1 = \frac{1}{2}$). Next, choose $a_2$ such that $2a_2 + a_1^2 = 0$ (thus $a_2 = -\frac{1}{8}$). Then choose $a_3$ such that the coefficient of $T^3$ on the right side equals 0 and so on so that the coefficient of $T^k$ (for $k \geq 2$) on the right side equals 0.

Actually, we do not care about the explicit formula for the $a_k$'s. We only need to know that some choice of the $a_k$'s gives a square root of $I + T$. ⊣

Note that every $z \in \mathbb{C}$ has a square root in $\mathbb{C}$. Write $z = r(\cos\theta + i \sin\theta)$ then $\sqrt{r}(\cos\frac{\theta}{2} + i \sin\frac{\theta}{2})$ is a square root of $z$.

**Theorem 8.26.** *Suppose $V$ is a complex vector space and $T \in \mathcal{L}(V)$ is invertible. Then $T$ has a square root.*

*Proof.* Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of $T$. For each $k$, there exists a nilpotent operator $T_k \in \mathcal{L}(G(\lambda_k, T))$ such that

$$T\big|_{G(\lambda_k,T)} = \lambda_k I + T_k.$$

Because $T$ is invertible, none of the $\lambda_k$'s equals 0, so we can write

$$T\big|_{G(\lambda_k,T)} = \lambda_k \left( I + \frac{T_k}{\lambda_k} \right)$$

for each $k$. Because $T_k/\lambda_k$ is nilpotent, $I + T_k/\lambda_k$ has a square root. Multiplying a square root of the complex number $\lambda_k$ by a square root of $I + T_k/\lambda_k$, we obtain a square root $R_k$ of $T\big|_{G(\lambda_k,T)}$. The construction of the square root of $R$ is then trivial. ⊣

**Definition 8.27.** Suppose $T \in \mathcal{L}(V)$. A basis of $V$ is called a *Jordan basis* for $T$ if, with respect to this basis, $T$ has a block diagonal matrix

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_p \end{pmatrix},$$

in which each $A_k$ is an upper–triangular matrix of the form

$$A_k = \begin{pmatrix} \lambda_k & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_k \end{pmatrix}.$$

**Theorem 8.28.** *Suppose $T \in \mathcal{L}(V)$ is nilpotent. Then there is a basis of $V$ that is a Jordan basis for $T$.*

*Proof.* We prove this result by induction on $\dim V$. To get started, note that the desired result holds if $\dim V = 1$ (because in that case, the only nilpotent operator is the 0 operator). Now assume that $\dim V > 1$ and that the desired result holds on all vector spaces of smaller dimension.

Let $m$ be the smallest positive integer such that $T^m = 0$. Thus there exists $u \in V$ such that $T^{m-1}u \neq 0$. Let

$$U = \operatorname{span}(u, Tu, \dots, T^{m-1}u).$$

By 8A.2 the list $u, Tu, \dots, T^{m-1}u$ is linearly independent. If $U = V$, then writing this list in reverse order gives a Jordan basis for $T$ and we are done. Thus we can assume that $U \neq V$.

Note that $U$ is invariant under $T$. By IH there is a basis of $U$ that is a Jordan basis for $T|_U$. Let $\varphi \in V'$ be such that $\varphi(T^{m-1}u) \neq 0$. Let

$$W = \{v \in V \,:\, \varphi(T^k v) = 0 \text{ for each } k = 0, \dots, m-1\}.$$

Then $W$ is a subspace of $V$ that is invariant under $T$.

We then show that $U + W$ is a direct sum. Suppose $v \in U \cap W$ with $v \neq 0$. Because $v \in U$, there exist $c_0, \dots, c_{m-1} \in \mathbb{F}$ such that

$$v = c_0 u + c_1 Tu + \cdots + c_{m-1} T^{m-1} u.$$

Let $j$ be the smallest index such that $c_j \neq 0$. Apply $T^{m-j-1}$ to both sides of the equation above, getting

$$T^{m-j-1} v = c_j T^{m-1} u.$$

Now apply $\varphi$ to both sides of the equation above, getting

$$\varphi(T^{m-j-1}v) = c_j\varphi(T^{m-1}u) \neq 0.$$

The equation above shows that $v \notin W$. Hence we have proved that $U \cap W = \{0\}$, which implies that $U + W$ is a direct sum.

To show that $U \oplus W = V$, define $S : V \to \mathbb{F}^m$ by

$$Sv = (\varphi(v), \varphi(Tv), \ldots, \varphi(T^{m-1}v)).$$

Thus null $S = W$. Hence

$$\dim W = \dim \operatorname{null} S = \dim V - \dim \operatorname{range} S \geq \dim V - m.$$

Using the inequality above, we have

$$\dim(U \oplus W) = \dim U + \dim W \geq m + (\dim V - m) = \dim V.$$

Thus $U \oplus W = V$. Again by our induction hypothesis, there will be a basis of $W$ that is a Jordan basis for $T|_W$. Putting together the Jordan bases for $T|_U$ and $T|_W$, we will have a Jordan basis for $T$. ⊣

**Theorem 8.29.** *Suppose $F = C$ and $T \in \mathcal{L}(V)$. Then there is a basis of $V$ that is a Jordan basis for $T$.*

*Proof.* Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of $T$. The generalized eigenspace decomposition states that

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T),$$

where each $(T - \lambda_k I)\big|_{G(\lambda_k,T)}$ is nilpotent. Thus 8.28 implies that some basis of each $G(\lambda_k, T)$ is a Jordan basis for $(T - \lambda_k I)\big|_{G(\lambda_k,T)}$. Put these bases together to get a basis of $V$ that is a Jordan basis for $T$. ⊣

## 8D Trace: A Connection Between Matrices and Operators

**Definition 8.30.** Suppose $A$ is a square matrix with entries in $\mathbb{F}$. The *trace* of $A$, denoted by $\operatorname{tr} A$, is defined to be the sum of the diagonal entries of $A$.

**Theorem 8.31.** *Suppose $A$ is an $m$-by-$n$ matrix and $B$ is an $n$-by-$m$ matrix. Then* $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

This is by computation.

**Theorem 8.32.** *Suppose $T \in \mathcal{L}(V)$. Suppose $u_1, \dots, u_n$ and $v_1, \dots, v_n$ are bases of $V$. Then*

$$\operatorname{tr} \mathcal{M}(T, (u_1, \dots, u_n)) = \operatorname{tr} \mathcal{M}(T, (v_1, \dots, v_n)).$$

*Proof.* Let $A = \mathcal{M}(T, (u_1, \dots, u_n))$ and $B = \mathcal{M}(T, (v_1, \dots, v_n))$. There exists an invertible $n \times n$ matrix $C$ such that $A = C^{-1}BC$ (see 3.84). Thus

$$\operatorname{tr} A = \operatorname{tr}\left((C^{-1}B)C\right) = \operatorname{tr}\left(C(C^{-1}B)\right) = \operatorname{tr}\left((CC^{-1})B\right) = \operatorname{tr} B. \qquad \dashv$$

**Definition 8.33.** Suppose $T \in \mathcal{L}(V)$. The *trace* of $T$, denoted $\operatorname{tr} T$, is defined by

$$\operatorname{tr} T = \operatorname{tr} \mathcal{M}(T, (v_1, \dots, v_n)),$$

where $v_1, \dots, v_n$ is any basis of $V$.

**Theorem 8.34.** *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Then $\operatorname{tr} T$ equals the sum of the eigenvalues of $T$, with each eigenvalue included as many times as its multiplicity.*

*Proof.* This is by 8.22. $\dashv$

**Theorem 8.35.** *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then $\operatorname{tr} T$ equals the negative of the coefficient of $z^{n-1}$ in the characteristic polynomial of $T$.*

**Theorem 8.36.** *Suppose $V$ is an inner product space, $T \in \mathcal{L}(V)$, and $e_1, \dots, e_n$ is an orthonormal basis of $V$. Then*

$$\operatorname{tr} T = \langle Te_1, e_1 \rangle + \cdots + \langle Te_n, e_n \rangle.$$

**Theorem 8.37.** *The function $\operatorname{tr} : \mathcal{L}(V) \to \mathbb{F}$ is a linear functional on $\mathcal{L}(V)$ such that*

$$\operatorname{tr}(ST) = \operatorname{tr}(TS)$$

*for all* $S, T \in \mathcal{L}(V)$.

**Theorem 8.38.** *There do not exist operators* $S, T \in \mathcal{L}(V)$ *such that* $ST - TS = I$.

*Proof.* Suppose $S, T \in \mathcal{L}(V)$. Then

$$\operatorname{tr}(ST - TS) = \operatorname{tr}(ST) - \operatorname{tr}(TS) = 0.$$

The trace of $I$ equals $\dim V$, which is not $0$. Because $ST - TS$ and $I$ have different traces, they cannot be equal. $\dashv$

Chapter 9

# Multilinear Algebra and Determinants

## 9A Bilinear Forms and Quadratic Forms

**Theorem 9.1.** *A* bilinear form *on $V$ is a function $\beta : V \times V \to \mathbb{F}$ such that*

$$v \mapsto \beta(v, u) \quad \text{and} \quad v \mapsto \beta(u, v)$$

*are both linear functionals on $V$ for every $u \in V$.*

The set of bilinear forms on $V$ is denoted by $V^{(2)}$.

**Theorem 9.2.** *Suppose $\beta$ is a bilinear form on $V$ and $e_1, \dots, e_n$ is a basis of $V$. The* matrix of $\beta$ *with respect to this basis is the $n$-by-$n$ matrix $\mathcal{M}(\beta)$ whose entry $\mathcal{M}(\beta)_{j,k}$ in row $j$, column $k$ is given by*

$$\mathcal{M}(\beta)_{j,k} = \beta(e_j, e_k).$$

If the basis is not clear, then the notation $\mathcal{M}(\beta, (e_1, \dots, e_n))$ is used.

**Theorem 9.3.** *Suppose $e_1, \dots, e_n$ is a basis of $V$. Then the map $\beta \mapsto \mathcal{M}(\beta)$ is an isomorphism of $V^{(2)}$ onto $\mathbb{F}^{n,n}$. Furthermore,* $\dim V^{(2)} = (\dim V)^2$.

*Proof Sketch.* Note that $\beta \mapsto \mathcal{M}(\beta)$ and $A \mapsto \beta_A$ are inverses of each other. ⊣

**Theorem 9.4.** *Suppose $\beta$ is a bilinear form on $V$ and $T \in \mathcal{L}(V)$. Define bilinear forms $\alpha$ and $\rho$ on $V$ by*

$$\alpha(u, v) = \beta(u, Tv) \quad \text{and} \quad \rho(u, v) = \beta(Tu, v).$$

*Let $e_1, \dots, e_n$ be a basis of $V$. Then*

$$\mathcal{M}(\alpha) = \mathcal{M}(\beta)\mathcal{M}(T) \quad \text{and} \quad \mathcal{M}(\rho) = \mathcal{M}(T)^{\top}\mathcal{M}(\beta).$$

*Proof.* We compute. If $j, k \in \{1, \dots, n\}$, then

$$\begin{aligned}\mathcal{M}(\alpha)_{j,k} &= \alpha(e_j, e_k)\\ &= \beta(e_j, Te_k)\\ &= \beta\left(e_j, \sum_{m=1}^{n} \mathcal{M}(T)_{m,k} e_m\right)\\ &= \sum_{m=1}^{n} \beta(e_j, e_m)\mathcal{M}(T)_{m,k}\\ &= (\mathcal{M}(\beta)\mathcal{M}(T))_{j,k}.\end{aligned}$$

Thus $\mathcal{M}(\alpha) = \mathcal{M}(\beta)\mathcal{M}(T)$. The proof that $\mathcal{M}(\rho) = \mathcal{M}(T)^\top \mathcal{M}(\beta)$ is similar. ⊣

**Theorem 9.5.** *Suppose $\beta \in V^{(2)}$. Suppose $e_1, \dots, e_n$ and $f_1, \dots, f_n$ are bases of $V$. Let*

$$A = \mathcal{M}(\beta, (e_1, \dots, e_n)) \quad \textit{and} \quad B = \mathcal{M}(\beta, (f_1, \dots, f_n))$$

*and*

$$C = \mathcal{M}(I, (e_1, \dots, e_n), (f_1, \dots, f_n)).$$

*Then*

$$A = C^\top BC.$$

*Proof.* Let $T$ denote the operator corresponding to $C$. Define bilinear forms $\alpha, \rho$ on $V$ by

$$\alpha(u, v) = \beta(u, Tv) \quad \text{and} \quad \rho(u, v) = \alpha(Tu, v) = \beta(Tu, Tv).$$

Then

$$\beta(e_j, e_k) = \beta(Tf_j, Tf_k) = \rho(f_j, f_k) \quad \text{for all } j, k \in \{1, \dots, n\}.$$

Thus

$$A = \mathcal{M}(\rho, (f_1, \dots, f_n)) = C^\top \mathcal{M}(\alpha, (f_1, \dots, f_n)) = C^\top BC.$$ ⊣

Cf. 3.46. Both are change-of-basis formulas.

**Definition 9.6.** A bilinear form $\rho \in V^{(2)}$ is called *symmetric* if

$$\rho(u, w) = \rho(w, u)$$

for all $u, w \in V$. The set of symmetric bilinear forms on $V$ is denoted by $V^{(2)}_{\text{sym}}$.

**Theorem 9.7.** *Suppose $\rho \in V^{(2)}$. Then the following are equivalent.*

*(a) $\rho$ is a symmetric bilinear form on $V$.*

*(b) $\mathcal{M}(\rho, (e_1, \dots, e_n))$ is a symmetric matrix for every basis $e_1, \dots, e_n$ of $V$.*

*(c) $\mathcal{M}(\rho, (e_1, \dots, e_n))$ is a symmetric matrix for some basis $e_1, \dots, e_n$ of $V$.*

*(d) $\mathcal{M}(\rho, (e_1, \dots, e_n))$ is a diagonal matrix for some basis $e_1, \dots, e_n$ of $V$.*

**Definition 9.8.** A bilinear form $\alpha \in V^{(2)}$ is called *alternating* if

$$\alpha(v, v) = 0$$

for all $v \in V$. The set of alternating bilinear forms on $V$ is denoted by $V^{(2)}_{\text{alt}}$.

**Definition 9.9.** For $\beta$ a bilinear form on $V$, define a function $q_\beta : V \to \mathbb{F}$ by

$$q_\beta(v) = \beta(v, v).$$

A function $q : V \to \mathbb{F}$ is called a *quadratic form* on $V$ if there exists a bilinear form $\beta$ on $V$ such that $q = q_\beta$.

## 9B Alternating Multilinear Forms

**Definition 9.10.** For $m$ a positive integer, an *$m$-linear form* on $V$ is a function

$$\beta : V^m \to \mathbb{F}$$

that is linear in each slot when the other slots are held fixed. The set of $m$-linear forms on $V$ is denoted by $V^{(m)}$. A function $\beta$ is called a *multilinear form* on $V$ if it is an $m$-linear form on $V$ for some positive integer $m$.

**Definition 9.11.** Suppose $m$ is a positive integer. An $m$-linear form $\alpha$ on $V$ is called *alternating* if

$$\alpha(v_1, \dots, v_m) = 0$$

whenever $v_1, \dots, v_m$ is a list of vectors in $V$ with $v_j = v_k$ for some two distinct values of $j$ and

$k$ in $\{1, \ldots, m\}$. The set of alternating $m$-linear forms on $V$ is denoted by $V_{\text{alt}}^{(m)}$.

**Theorem 9.12.** *Suppose $m$ is a positive integer and $\alpha$ is an alternating $m$-linear form on $V$. If $v_1, \ldots, v_m$ is a linearly dependent list in $V$, then*

$$\alpha(v_1, \ldots, v_m) = 0.$$

**Theorem 9.13.** *For $m > \dim V$, $0$ is the only alternating $m$-linear form on $V$.*

**Theorem 9.14.** *Suppose $m$ is a positive integer, $\alpha$ is an alternating $m$-linear form on $V$, and $v_1, \ldots, v_m$ is a list of vectors in $V$. Then swapping the vectors in any two slots of $\alpha(v_1, \ldots, v_m)$ changes the value of $\alpha$ by a factor of $-1$.*

*Proof.* We prove for the case $m = 2$.

$$\begin{aligned} &\alpha(v_1, v_2) + \alpha(v_2, v_1) \\ &= \alpha(\frac{v_1 + v_2}{2} + \frac{v_1 - v_2}{2}, \frac{v_1 + v_2}{2} - \frac{v_1 - v_2}{2}) \\ &\quad + \alpha(\frac{v_1 + v_2}{2} - \frac{v_1 - v_2}{2}, \frac{v_1 + v_2}{2} + \frac{v_1 - v_2}{2}) \\ &= 0 \end{aligned}$$

⊣

**Definition 9.15.** Suppose $m$ is a positive integer. A *permutation* of $(1, \ldots, m)$ is a list $(j_1, \ldots, j_m)$ that contains each of the numbers $1, \ldots, m$ exactly once. The set of all permutations of $(1, \ldots, m)$ is denoted by $\operatorname{perm} m$.

**Definition 9.16.** The *sign* of a permutation $(j_1, \ldots, j_m)$ is defined by

$$\operatorname{sign}(j_1, \ldots, j_m) = (-1)^N,$$

where $N$ is the number of pairs of integers $(k, \ell)$ with $1 \leq k < \ell \leq m$ such that $k$ appears after $\ell$ in the list $(j_1, \ldots, j_m)$.

**Theorem 9.17.** *Swapping two entries in a permutation multiplies the sign of the permutation by* $-1$.

**Theorem 9.18.** *Suppose* $m$ *is a positive integer and* $\alpha \in V_{\text{alt}}^{(m)}$. *Then*

$$\alpha(v_{j_1}, \dots, v_{j_m}) = \operatorname{sign}(j_1, \dots, j_m)\, \alpha(v_1, \dots, v_m)$$

*for every list* $v_1, \dots, v_m$ *of vectors in* $V$ *and all* $(j_1, \dots, j_m) \in \operatorname{perm} m$.

**Theorem 9.19.** *Let* $n = \dim V$. *Suppose* $e_1, \dots, e_n$ *is a basis of* $V$ *and* $v_1, \dots, v_n \in V$. *For each* $k \in \{1, \dots, n\}$, *let* $b_{1,k}, \dots, b_{n,k} \in \mathbb{F}$ *be such that*

$$v_k = \sum_{j=1}^{n} b_{j,k} e_j.$$

*Then*

$$\alpha(v_1, \dots, v_n) = \alpha(e_1, \dots, e_n) \sum_{(j_1,\dots,j_n)\in \operatorname{perm} n} (\operatorname{sign}(j_1, \dots, j_n)) b_{j_1,1} \cdots b_{j_n,n},$$

*for every alternating* $n$*-linear form* $\alpha$ *on* $V$.

*Proof.* Suppose $\alpha$ is an alternating $n$-linear form $\alpha$ on $V$. Then

$$\begin{aligned}
\alpha(v_1, \dots, v_n) &= \alpha\left(\sum_{j_1=1}^{n} b_{j_1,1} e_{j_1}, \ \dots, \ \sum_{j_n=1}^{n} b_{j_n,n} e_{j_n}\right) \\
&= \sum_{j_1=1}^{n} \cdots \sum_{j_n=1}^{n} b_{j_1,1} \cdots b_{j_n,n}\, \alpha(e_{j_1}, \dots, e_{j_n}) \\
&= \sum_{(j_1,\dots,j_n)\in \operatorname{perm}_n} b_{j_1,1} \cdots b_{j_n,n}\, \alpha(e_{j_1}, \dots, e_{j_n}) \\
&= \alpha(e_1, \dots, e_n) \sum_{(j_1,\dots,j_n)\in \operatorname{perm}_n} (\operatorname{sign}(j_1, \dots, j_n))\, b_{j_1,1} \cdots b_{j_n,n},
\end{aligned}$$

where the third line holds because $\alpha(e_{j_1}, \dots, e_{j_n}) = 0$ if $j_1, \dots, j_n$ are not distinct integers. ⊣

**Theorem 9.20.** *The vector space* $V_{\text{alt}}^{(\dim V)}$ *has dimension one.*

*Proof.* Let $n = \dim V$. Suppose $\alpha$ and $\alpha'$ are alternating $n$-linear forms on $V$ with $\alpha \neq 0$. Let $e_1, \dots, e_n$ be such that $\alpha(e_1, \dots, e_n) \neq 0$. There exists $c \in \mathbb{F}$ such that

$$\alpha'(e_1, \dots, e_n) = c\,\alpha(e_1, \dots, e_n).$$

9.12 implies that $e_1, \dots, e_n$ is linearly independent and thus a basis of $V$.

Suppose $v_1, \dots, v_n \in V$. Let $b_{j,k}$ be as in 9.19 for $j, k = 1, \dots, n$. Then

$$\begin{aligned}
\alpha'(v_1, \dots, v_n) &= \alpha'(e_1, \dots, e_n) \sum_{(j_1,\dots,j_n)\in\text{perm}\,n} (\text{sign}(j_1, \dots, j_n)) b_{j_1,1} \cdots b_{j_n,n} \\
&= c\,\alpha(e_1, \dots, e_n) \sum_{(j_1,\dots,j_n)\in\text{perm}\,n} (\text{sign}(j_1, \dots, j_n)) b_{j_1,1} \cdots b_{j_n,n} \\
&= c\,\alpha(v_1, \dots, v_n),
\end{aligned}$$

The equation above implies that $\dim V_{\text{alt}}^{(n)} \leq 1$.

To complete the proof, we eliminate the possibility that $\dim V_{\text{alt}}^{(n)} = 0$. Let $e_1, \dots, e_n$ be any basis of $V$, and let $\varphi_1, \dots, \varphi_n \in V'$ be the linear functionals on $V$ that

$$v = \sum_{j=1}^{n} \varphi_j(v) e_j$$

for every $v \in V$. Now for $v_1, \dots, v_n \in V$, define

$$\alpha(v_1, \dots, v_n) = \sum_{(j_1,\dots,j_n)\in\text{perm}\,n} (\text{sign}(j_1, \dots, j_n)) \varphi_{j_1}(v_1) \cdots \varphi_{j_n}(v_n). \tag{$*$}$$

The verification that $\alpha$ is an $n$-linear form on $V$ is straightforward.

Suppose $v_1, \dots, v_n \in V$ with $v_1 = v_2$. For each $(j_1, \dots, j_n) \in \text{perm}\,n$, the permutation $(j_2, j_1, j_3, \dots, j_n)$ has the opposite sign. Because $v_1 = v_2$, the contributions from these two permutations to the sum in $(*)$ cancel each other. Hence $\alpha(v_1, v_1, v_3, \dots, v_n) = 0$. Similarly, $\alpha(v_1, \dots, v_n) = 0$ if any two vectors in the list $v_1, \dots, v_n$ are equal. Thus $\alpha$ is alternating.

Finally, consider $(*)$ with each $v_k = e_k$. Because $\varphi_j(e_k)$ equals 0 if $j \neq k$ and equals 1 if $j = k$, only the permutation $(1, \dots, n)$ makes a nonzero contribution to the right side of $(*)$ in this case, giving the equation $\alpha(e_1, \dots, e_n) = 1$. Thus we have produced a nonzero alternating $n$-linear form $\alpha$ on $V$, as desired. $\dashv$

**Theorem 9.21.** *Let $n = \dim V$. Suppose $\alpha$ is a nonzero alternating $n$-linear form on $V$ and*

*$e_1, \ldots, e_n$ is a list of vectors in $V$. Then*

$$\alpha(e_1, \ldots, e_n) \neq 0$$

*if and only if $e_1, \ldots, e_n$ is linearly independent.*

*Proof.* First suppose $\alpha(e_1, \ldots, e_n) \neq 0$. Then 9.12 implies that $e_1, \ldots, e_n$ is linearly independent.

Now suppose $e_1, \ldots, e_n$ is linearly independent. Because $n = \dim V$, this implies that $e_1, \ldots, e_n$ is a basis of $V$.

Because $\alpha$ is not the zero $n$-linear form, there exist $v_1, \ldots, v_n \in V$ such that $\alpha(v_1, \ldots, v_n) \neq 0$. Now 9.19 implies that $\alpha(e_1, \ldots, e_n) \neq 0$. ⊣

## 9C Determinants

**Definition 9.22.** Suppose that $m$ is a positive integer and $T \in \mathcal{L}(V)$. For $\alpha \in V_{\text{alt}}^{(m)}$, define $\alpha_T \in V_{\text{alt}}^{(m)}$ by

$$\alpha_T(v_1, \ldots, v_m) = \alpha(Tv_1, \ldots, Tv_m)$$

for each list $v_1, \ldots, v_m$ of vectors in $V$.

Note that the function $\alpha \mapsto \alpha_T$ is a linear map of $V_{\text{alt}}^{(m)}$ to itself.

**Definition 9.23.** Suppose $T \in \mathcal{L}(V)$. The *determinant* of $T$, denoted by $\det T$, is defined to be the unique number in $\mathbb{F}$ such that

$$\alpha_T = (\det T)\, \alpha$$

for all $\alpha \in V_{\text{alt}}^{(\dim V)}$.

**Definition 9.24.** Suppose that $n$ is a positive integer and $A$ is an $n$-by-$n$ square matrix with entries in $\mathbb{F}$. Let $T \in \mathcal{L}(\mathbb{F}^n)$ be the operator whose matrix with respect to the standard basis of $\mathbb{F}^n$ equals $A$. The *determinant* of $A$, denoted by $\det A$, is defined by $\det A = \det T$.

**Theorem 9.25.** *Suppose that $n$ is a positive integer. The map that takes a list $v_1, \ldots, v_n$ of vectors in $\mathbb{F}^n$ to $\det(v_1 \ \cdots \ v_n)$ is an alternating $n$-linear form on $\mathbb{F}^n$.*

*Proof.* Let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{F}^n$ and suppose $v_1, \ldots, v_n$ is a list of vectors in $\mathbb{F}^n$. Let $T \in \mathcal{L}(\mathbb{F}^n)$ be the operator such that $Te_k = v_k$ for $k = 1, \ldots, n$. Thus $T$ is the operator whose matrix with respect to $e_1, \ldots, e_n$ is $(\, v_1 \ \cdots \ v_n\,)$. Hence $\det(v_1 \ \cdots \ v_n) = \det T$ by definition of the determinant of a matrix.

Let $\alpha$ be an alternating $n$-linear form on $\mathbb{F}^n$ such that $\alpha(e_1, \ldots, e_n) = 1$. Then

$$\begin{aligned} \det(v_1 \ \cdots \ v_n) &= \det T \\ &= (\det T)\, \alpha(e_1, \ldots, e_n) \\ &= \alpha(Te_1, \ldots, Te_n) \\ &= \alpha(v_1, \ldots, v_n), \end{aligned}$$

as desired. ⊣

**Theorem 9.26.** *Suppose that $n$ is a positive integer and $A$ is an $n$-by-$n$ square matrix. Then*

$$\det A = \sum_{(j_1,\ldots,j_n)\in \operatorname{perm} n} \big(\operatorname{sign}(j_1, \ldots, j_n)\big)\, A_{j_1,1} \cdots A_{j_n,n}.$$

This is trivial by 9.19 and 9.25.

**Theorem 9.27.** *Suppose that $A$ is an upper-triangular matrix with $\lambda_1, \ldots, \lambda_n$ on the diagonal. Then*

$$\det A = \lambda_1 \cdots \lambda_n.$$

*Proof.* If $(j_1, \ldots, j_n) \in \operatorname{perm} n$ with $(j_1, \ldots, j_n) \neq (1, \ldots, n)$, then $j_k > k$ for some $k \in \{1, \ldots, n\}$, which implies that $A_{j_k,k} = 0$. Because $A_{k,k} = \lambda_k$ for each $k = 1, \ldots, n$, this implies that

$$\det A = \lambda_1 \cdots \lambda_n.$$

⊣

**Theorem 9.28.** *Suppose $S, T \in \mathcal{L}(V)$. Then $\det(ST) = (\det S)(\det T)$. Similarly for $A$ and $B$ square matrices of the same size.*

**Theorem 9.29.** *An operator $T \in \mathcal{L}(V)$ is invertible if and only if* $\det T \neq 0$. *Furthermore, if $T$ is invertible, then*

$$\det(T^{-1}) = \frac{1}{\det T}.$$

*Proof.* First suppose $T$ is invertible. Thus $TT^{-1} = I$, and

$$1 = \det I = \det(TT^{-1}) = (\det T)(\det(T^{-1})).$$

Hence $\det T \neq 0$ and $\det(T^{-1})$ is the multiplicative inverse of $\det T$.

To prove the other direction, now suppose $\det T \neq 0$. Suppose $v \in V$ and $v \neq 0$. Let $v, e_2, \dots, e_n$ be a basis of $V$ and let $\alpha \in V_{\text{alt}}^{(n)}$ be such that $\alpha \neq 0$. Then $\alpha(v, e_2, \dots, e_n) \neq 0$ (by 9.21). Now

$$\alpha(Tv, Te_2, \dots, Te_n) = (\det T)\,\alpha(v, e_2, \dots, e_n) \neq 0.$$

Thus $Tv \neq 0$. Hence $T$ is invertible. ⊣

**Theorem 9.30.** *Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. Then $\lambda$ is an eigenvalue of $T$ if and only if*

$$\det(\lambda I - T) = 0.$$

**Theorem 9.31.** *Suppose $T \in \mathcal{L}(V)$ and $S : W \to V$ is an invertible linear map. Then*

$$\det(S^{-1}TS) = \det T.$$

*Proof.* Let $n = \dim W = \dim V$. Suppose $\tau \in W_{\text{alt}}^{(n)}$. Define $\alpha \in V_{\text{alt}}^{(n)}$ by

$$\alpha(v_1, \dots, v_n) = \tau(S^{-1}v_1, \dots, S^{-1}v_n)$$

for $v_1, \dots, v_n \in V$. Suppose $w_1, \dots, w_n \in W$. Then

$$\begin{aligned}\tau_{S^{-1}TS}(w_1, \dots, w_n) &= \tau(S^{-1}TSw_1, \dots, S^{-1}TSw_n)\\ &= \alpha(TSw_1, \dots, TSw_n)\\ &= (\det T)\alpha(Sw_1, \dots, Sw_n)\\ &= (\det T)\tau(w_1, \dots, w_n).\end{aligned}$$

The equation above and the definition of the determinant of the operator $S^{-1}TS$ imply that $\det(S^{-1}TS) = \det T$. ⊣

**Theorem 9.32.** *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Then $\det T$ equals the product of the eigenvalues of $T$, with each eigenvalue included as many times as its multiplicity.*

This is by 8.24.

**Theorem 9.33.** *Determinant of transpose, dual, or adjoint.*

*(a) Suppose $A$ is a square matrix. Then $\det A^{\mathrm{t}} = \det A$.*
*(b) Suppose $T \in \mathcal{L}(V)$. Then $\det T' = \det T$.*
*(c) Suppose $V$ is an inner product space and $T \in \mathcal{L}(V)$. Then $\det(T^*) = \overline{\det T}$.*

**Theorem 9.34.** *Suppose $V$ is an inner product space and $S \in \mathcal{L}(V)$ is a unitary operator. Then $|\det S| = 1$.*

*Proof.* Because $S$ is unitary, $I = S^*S$. Thus

$$1 = \det(S^*S) = (\det S^*)(\det S) = \overline{(\det S)}\,(\det S) = |\det S|^2. \qquad \dashv$$

**Theorem 9.35.** *Suppose $V$ is an inner product space and $T \in \mathcal{L}(V)$ is a positive operator. Then $\det T \geq 0$.*

**Theorem 9.36.** *Suppose $V$ is an inner product space and $T \in \mathcal{L}(V)$. Then*

$$|\det T| = \sqrt{\det(T^*T)} = \textit{product of singular values of } T.$$

**Theorem 9.37.** *Suppose $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of $T$, and let $d_1, \ldots, d_m$ denote their multiplicities. Then*

$$\det(zI - T) = (z - \lambda_1)^{d_1} \cdots (z - \lambda_m)^{d_m}.$$

A new definition that is equivalent to 8.18 when $\mathbb{F} = \mathbb{C}$.

**Definition 9.38.** Suppose $T \in \mathcal{L}(V)$. The polynomial defined by

$$z \mapsto \det(zI - T)$$

is called the *characteristic polynomial* of $T$.

**Theorem 9.39.** ***Cayley–Hamilton, extended*** *Suppose $T \in \mathcal{L}(V)$ and $q$ is the characteristic polynomial of $T$. Then $q(T) = 0$.*

*Proof.* If $\mathbb{F} = \mathbb{C}$, then the equation $q(T) = 0$ follows from 8.20.

Now suppose $\mathbb{F} = \mathbb{R}$. Fix a basis of $V$, and let $A$ be the matrix of $T$ with respect to this basis. Let $S$ be the operator on $\mathbb{C}^{\dim V}$ such that the matrix of $S$ (with respect to the standard basis of $\mathbb{C}^{\dim V}$) is $A$. For all $z \in \mathbb{R}$ we have

$$q(z) = \det(zI - T) = \det(zI - A) = \det(zI - S).$$

Thus $q$ is the characteristic polynomial of $S$. The case $\mathbb{F} = \mathbb{C}$ (first sentence of this proof) now implies that

$$0 = q(S) = q(A) = q(T). \qquad \dashv$$