Linear Algebra Done Right Notes

Steven Labalme

October 27, 2021

Contents

List of Figures

Chapter 1

Vector Spaces

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

10/27:

- Assumed familiarity with the set $\mathbb{R}$ of real numbers.
- **Complex number**: An ordered pair (a, b), where $a, b \in \mathbb{R}$, but we will write this as $a + bi$.
  - The set of all complex number is denoted by $\mathbb{C}$:

$$\mathbb{C} = \{a + bi : a, b \in \mathbb{R}\}^{[1]}$$

  - Definitions of **addition** and **multiplication** on $\mathbb{C}$ are given, but I know these.
- Properties of complex arithmetic:
  - **Commutativity**: $\alpha + \beta = \beta + \alpha$ and $\alpha\beta = \beta\alpha$ for all $\alpha, \beta \in \mathbb{C}$.
  - **Associativity**: $(\alpha + \beta) + \lambda = \alpha + (\beta + \lambda)$ and $(\alpha\beta)\lambda = \alpha(\beta\lambda)$ for all $\alpha, \beta, \lambda \in \mathbb{C}$.
  - **Identities**: $\lambda + 0 = \lambda$ and $\lambda 1 = \lambda$ for all $\lambda \in \mathbb{C}$.
  - **Additive inverse**: For every $\alpha \in \mathbb{C}$, there exists a unique $\beta \in \mathbb{C}$ such that $\alpha + \beta = 0$.
  - **Multiplicative inverse**: For every $\alpha \in \mathbb{C}$ with $\alpha \neq 0$, there exists a unique $\beta \in \mathbb{C}$ such that $\alpha\beta = 1$.
  - **Distributive property**: $\lambda(\alpha + \beta) = \lambda\alpha + \lambda\beta$ for all $\lambda, \alpha, \beta \in \mathbb{C}$.
- "The properties above are proved using the familiar properties of real numbers and the definitions of complex addition and multiplication" (Axler, 2015, p. 3).
- $\mathbb{F}$ stands for $\mathbb{R}$ or $\mathbb{C}$.
  - Any theorem proved with $\mathbb{F}$ holds when $\mathbb{F}$ is replaced with $\mathbb{R}$ and when $\mathbb{F}$ is replaced with $\mathbb{C}$.
- **Scalar**: A number or magnitude. This word is commonly used to differentiate a quantity from a **vector** quantity.
- Subtraction and division are defined.
- Properties of exponents are defined.
- The set $\mathbb{R}^2$, which can be conceived as a plane, is the set of all **ordered pairs** of real numbers:

$$\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}$$

[1] The complex numbers equal the set of numbers $a + bi$ such that a and b are elements of the real numbers.

- The set $\mathbb{R}^3$, which can be conceived as ordinary space, is the set of all **ordered triples** of real numbers:

$$\mathbb{R}^3 = \{(x, y, z) : x, y, z \in \mathbb{R}\}$$

- "Suppose n is a nonnegative integer. A **list** of **length** n is an ordered collection of n elements (which might be numbers, other lists, or more abstract entities) separated by commas and surrounded by parentheses. A list of length n looks like this:

$$(x_1, \dots, x_n)$$

  Two lists are equal if and only if they have the same length and the same elements in the same order" (Axler, 2015, p. 5).

- **Ordered pair**: A list of length 2.
- **Ordered triple**: A list of length 3.
- ***n*-tuple**: A list of length n.
- Although lists are sometimes discussed without specifying their length, a list must, by definition, have a finite length, i.e. $(x_1, x_2, \dots)$ is not a list.
- A list of length 0 looks like this: $()$.
  - Such an object is defined to avoid trivial exceptions to theorems.
- Lists vs. **sets**: In lists, order matters and repetitions have meaning. In sets, order and repetitions are irrelevant.
- "$\mathbb{F}^n$ is the set of all lists of length n of elements of $\mathbb{F}$:

$$\mathbb{F}^n = \{(x_1, \dots, x_n) : x_j \in \mathbb{F} \text{ for } j = 1, \dots, n)\}$$

  For $(x_1, \dots, x_n) \in \mathbb{F}^n$ and $j \in \{1, \dots, n\}$, we say that x_j is the j^{th} **coordinate** of $(x_1, \dots, x_n)$" (Axler, 2015, p. 6).

- For help in conceiving higher dimensional spaces, consider reading Abbott (1952). This is an amusing account of how $\mathbb{R}^3$ would be perceived by creatures living in $\mathbb{R}^2$.
- **Addition** (in $\mathbb{F}^n$): Add corresponding coordinates:

$$(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n)$$

- For a simpler notation, use a single letter to denote a list of n numbers.
  - **Commutativity** (of addition in $\mathbb{F}^n$): If $x, y \in \mathbb{F}^n$, then $x + y = y + x$.
  - However, the proof still requires the more formal, cumbersome list notation.
- **0**: The list of length n whose coordinates are all 0:

$$0 = (0, \dots, 0)$$

  - Although the ambiguity in the use of "0" on the left vs. right side of the equation may seem confusing, context can always differentiate between which definition is needed.
- A picture can help visualize $\mathbb{R}^2$ because $\mathbb{R}^2$ can be sketched on 2-dimensional surfaces such as paper.
  - A typical element of $\mathbb{R}^2$ is a point $x = (x_1, x_2)$.
  - However, points are generally though of as an arrow starting at the origin and ending at x, as shown below.

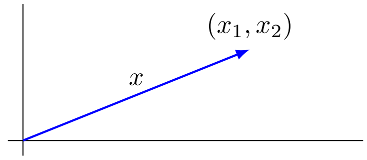

Figure 1.1: $x \in \mathbb{R}^2$ can be conceived as a point or a vector.

- When thought of as an arrow, x is called a **vector**.
- When translated without varying length or direction, it is still the same vector.
- Remember that these pictures are aids — although we cannot visualize higher dimensional vector spaces, the algebraic elements are as rigorously defined as those of $\mathbb{R}^2$.

- Addition has a simple geometric interpretation in $\mathbb{R}^2$.
- If we want to add $x + y$, slide y so that its initial point coincides with the terminal point of x. The sum is the vector from the tail of x to the head of y.

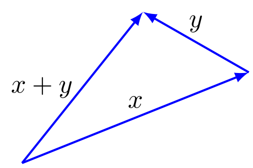

Figure 1.2: Vector addition.

- "For $x \in \mathbb{F}^n$, the **additive inverse** of x, denoted $-x$, is the vector $-x \in \mathbb{F}^n$ such that

$$x + (-x) = 0$$

In other words, if $x = (x_1, \dots, x_n)$, then $-x = (-x_1, \dots, -x_n)$" (Axler, 2015, p. 9).

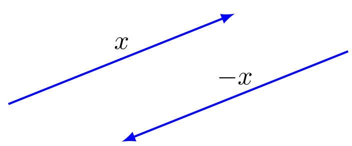

Figure 1.3: A vector and its additive inverse.

  - For $x \in \mathbb{R}^2$, $-x$ is the vector parallel to x with the same length but in the opposite direction.

- **Product (scalar multiplication)**: When multiplying $\lambda \in \mathbb{F}$ and $x \in \mathbb{F}^n$, multiply each coordinate of x by λ:

$$\lambda (x_1, \dots, x_n) = (\lambda x_1, \dots, \lambda x_n)$$

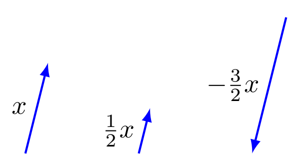

Figure 1.4: Scalar multiplication.

- **Field**: A "set containing at least two distinct elements called 0 and 1, along with operations of addition and multiplication satisfying all the properties" of complex arithmetic (see earlier in this section) (Axler, 2015, p. 10).

1.B Definition of Vector Space

- **Addition (on a set V)**: "A function that assigns an element $u + v \in V$ to each pair of elements $u, v \in V$" (Axler, 2015, p. 12).
- **Scalar multiplication (on a set V)**: "A function that assigns an element $\lambda v \in V$ to each $\lambda \in \mathbb{F}$ and each $v \in V$" (Axler, 2015, p. 12).
- **Vector space**: "A set V along with an addition and a scalar multiplication on V such that the following properties hold:" (Axler, 2015, p. 12).

  commutativity
  $u + v = v + u$ for all $u, v \in V$

  associativity
  $(u + v) + w = u + (v + w)$ and $(ab)v = a(bv)$ for all $u, v, w \in V$ and all $a, b \in \mathbb{F}$

  additive identity
  There exists an element $0 \in V$ such that $v + 0 = v$ for all $v \in V$

  additive inverse
  For every $v \in V$, there exists $w \in V$ such that $v + w = 0$

  multiplicative identity
  $1v = v$ for all $v \in V$

  distributive properties
  $a(u + v) = au + av$ and $(a + b)v = av + bv$ for all $a, b \in \mathbb{F}$ and all $u, v \in V$

- To be more precise, V depends on $\mathbb{F}$, so sometimes we say V is a **vector space over** $\mathbb{F}$.
  - For example, $\mathbb{R}^n$ is only a vector space over $\mathbb{R}$, not $\mathbb{C}$.
- **Real vector space**: A vector space over $\mathbb{R}$.
- **Complex vector space**: A vector space over $\mathbb{C}$.
- $\mathbb{F}^\infty$ is a vector space.
- $\mathbb{F}^S$ denotes the set of functions from S to $\mathbb{F}$.
  - For example, $\mathbb{R}^{[0,1]}$ is the "set of real-valued functions on the interval $[0, 1]$" (Axler, 2015, p. 14).
  - You can think of $\mathbb{F}^n$ as $\mathbb{F}^{\{1,2,\dots,n\}}$.
- Elementary properties of vector spaces:

  Theorem 1.1 (Unique additive identity). *A vector space has a unique additive identity.*

  Proof. Suppose 0 and $0'$ are both additive identities in V. Then

  $$0' = 0' + 0 = 0 + 0' = 0$$

  The first equality holds due to 0 being an additive identity. The second holds due to commutativity. The third holds due to $0'$ being an additive identity. Thus, $0 = 0'$, and V has only one additive identity. ∎

  Theorem 1.2 (Unique additive inverse). *Each element $v \in V$ has a unique additive inverse.*

Proof. Same idea:
$$w = w + 0 = w + (v + w') = (w + v) + w' = 0 + w' = w'$$
■

Theorem 1.3 (The number 0 times a vector)**.** *$0v = 0\ \forall\ v \in V$, where 0 on the left side is a scalar and 0 on the right side is a vector (the additive identity of V).*

Proof. Since this property asserts something about both scalar multiplication and the additive identity, the distributive property (the only part of the definition of a vector space that connects scalar multiplication and vector addition) must be used in the proof.
$$\begin{aligned} 0v &= (0+0)v \\ 0v &= 0v + 0v \\ 0v - 0v &= 0v + 0v - 0v \\ 0 &= 0v \end{aligned}$$
■

Theorem 1.4 (A number times the vector 0)**.** *$a0 = 0\ \forall\ a \in \mathbb{F}$, where 0 is a vector.*

Proof. Same as above. ■

Theorem 1.5 (The number -1 times a vector)**.** *$(-1)v = -v\ \forall\ v \in V$, where -1 is a scalar and $-v$ is the additive inverse of v.*

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

1.C Subspaces

- **Subspace**: A subset U of V that is a vector space under the same definition of addition and scalar multiplication as on V, e.g., satisfies the following three conditions.

  additive identity
  $0 \in U$

  closed under addition
  $u, w \in U$ implies $u + w \in U$

  closed under scalar multiplication
  $a \in \mathbb{F}$ and $u \in U$ implies $au \in U$

- The other conditions can be derived from the above 3.
- When we look at subspaces within the differentiable functions, the logical foundation of calculus appears.
- The subspaces of $\mathbb{R}^2$ are $\{0\}$, $\mathbb{R}^2$, and any straight line through the origin.
- The subspaces of $\mathbb{R}^3$ are $\{0\}$, $\mathbb{R}^3$, any straight line through the origin, and any flat plane through the origin.
- **Sum of subsets**: If $U_1, \dots, U_n$ are subsets of V, their **sum** (denoted $U_1 + \cdots + U_n$) is the set of all possible sums of elements of $U_1, \dots, U_n$:
$$U_1 + \cdots + U_m = \{u_1 + \cdots + u_m : u_1 \in U_1, \dots, u_m \in U_m\}$$

- The sum of subspaces is the smallest containing subspace.
  - Clearly, the sum of subspaces is a subspace (satisfies 3 tenets).
  - The sum of subspaces contains every original element (u_1 plus the 0 from u_2, etc.). Any subspace containing all of these elements must contain every finite sum of them (by definition). Thus, no smaller subspace can be created than that of the sum of every element.
- **Direct sum**: A sum of subspaces where each element of $U_1 + \cdots + U_m$ can be written in only one way as a sum $u_1 + \cdots + u_m$.
  - $U_1 \oplus \cdots \oplus U_m$ denotes $U_1 + \cdots + U_m$ if $U_1 + \cdots + U_m$ is a direct sum.
- A sum of subspaces is a direct sum if and only if the only way to write 0 as a sum of elements is by summing the 0 of each subset.
- A sum of subspaces U and W is a direct sum if and only if $U \cap W = \{0\}$.

Chapter 2

Finite-Dimensional Vector Spaces

2.A Span and Linear Independence

9/3:

- **Linear combination** (of a list $v_1, \dots, v_m$ of vectors in V): A vector of the form $a_1v_1 + \cdots + a_mv_m$, where $a_1, \dots, a_m \in \mathbb{F}$.
- **Span** (of $v_1, \dots, v_m$): The set of all linear combinations of a list of vectors $v_1, \dots, v_m$ in V. *Also known as* **linear span**. *Denoted by* $\text{span}(v_1, \dots, v_m)$. *Given by*

$$\text{span}(v_1, \dots, v_m) = \{a_1v_1 + \cdots + a_mv_m : a_1, \dots, a_m \in \mathbb{F}\}$$

  - We define $\text{span}() = \{0\}$.
- Span as a subspace.

Theorem 2.1. *The span of a list of vectors in V is the smallest subspace of V containing all the vectors in the list.*

Proof. Let $v_1, \dots, v_m \in V$ be a list of vectors. We will first prove that $\text{span}(v_1, \dots, v_m)$ is a subspace of V. We will then prove that it contains all of the vectors in the list. Lastly, we will prove that any proper subset of $\text{span}(v_1, \dots, v_m)$ either doesn't contain all the vectors in the list or is not a subspace of V. Let's begin.

To prove that $\text{span}(v_1, \dots, v_m)$ is a subspace of V, it will suffice to show that $\text{span}(v_1, \dots, v_m)$ contains the additive identity, $\text{span}(v_1, \dots, v_m)$ is closed under addition, and $\text{span}(v_1, \dots, v_m)$ is closed under scalar multiplication. By the definition of $\text{span}(v_1, \dots, v_m)$, we know that $0v_1 + \cdots + 0v_m = 0 \in \text{span}(v_1, \dots, v_m)$. If $a_1v_1 + \cdots + a_mv_m \in \text{span}(v_1, \dots, v_m)$ and $b_1v_1 + \cdots + b_mv_m \in \text{span}(v_1, \dots, v_m)$, then naturally $(a_1v_1 + \cdots + a_mv_m) + (b_1v_1 + \cdots + b_mv_m) = (a_1 + b_1)v_1 + \cdots + (a_m + b_m)v_m \in \text{span}(v_1, \dots, v_m)$. Lastly, if $a_1v_1 + \cdots + a_mv_m \in \text{span}(v_1, \dots, v_m)$ and $\lambda \in \mathbb{F}$, then naturally $\lambda(a_1v_1 + \cdots + a_mv_m) = (\lambda a_1)v_1 + \cdots + (\lambda a_m)v_m \in \text{span}(v_1, \dots, v_m)$.

By setting every $a_i = 0$ except $a_j = 1$, we can guarantee that $v_j \in \text{span}(v_1, \dots, v_m)$ for all $j \in [m]$.

Suppose for the sake of contradiction that there exists a smaller subspace U of V that contains $v_1, \dots, v_m$. It follows that there exists a vector $u \in \text{span}(v_1, \dots, v_m)$ such that $u \notin U$. Since $u \in \text{span}(v_1, \dots, v_m)$, $u = a_1v_1 + \cdots + a_mv_m$ for some $a_1, \dots, a_m \in \mathbb{F}$. However, by definition, $v_1, \dots, v_m \in U$, so since U is closed under addition and scalar multiplication, their linear combination $a_1v_1 + \cdots + a_mv_m = u \in U$, a contradiction. ■

- If $\text{span}(v_1, \dots, v_m) = V$, we say that $v_1, \dots, v_m$ **spans** V.
- **Finite-dimensional vector space**: A vector space such that some list of vectors in it spans the space.

  - Recall that by definition, every list has finite length.

- **Polynomial** (with coefficients in $\mathbb{F}$): A function $p : \mathbb{F} \to \mathbb{F}$ such that 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 $z \in \mathbb{F}$.

- $\boldsymbol{\mathcal{P}(\mathbb{F})}$: The set of all polynomials with coefficients in $\mathbb{F}$.
  - $\mathcal{P}(\mathbb{F})$, under the usual addition and scalar multiplication, is a vector space over $\mathbb{F}$.
  - Thus, $\mathcal{P}(\mathbb{F})$ is a subspace of $\mathbb{F}^{\mathbb{F}}$.

- We will later prove that the coefficients of a polynomial uniquely determine it.

- **Degree** (of a polynomial p): The number m, where $p = a_0 + a_1 z + \cdots + a_m z^m$ and $a_m \neq 0$. *Denoted by* $\deg p = m$.
  - The polynomial $p(z) = 0$ is said to have degree $-\infty$.

- $\boldsymbol{\mathcal{P}_m(\mathbb{F})}$: The set of all polynomials with coefficients in $\mathbb{F}$ and degree at most m, where m is a nonnegative integer.
  - $\mathcal{P}_m(\mathbb{F}) = \operatorname{span}(1, z, \dots, z^m)$ is a finite-dimensional vector space for all nonnegative integers m.

- **Infinite-dimensional vector space**: A vector space that is not finite dimensional.
  - $\mathcal{P}(\mathbb{F})$ is infinite-dimensional.

- **Linearly independent** (list $v_1, \dots, v_m$): A list $v_1, \dots, v_m$ of vectors in V such that 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 $a_1 = \cdots = a_m = 0$.
  - We also let the empty list be linearly independent.

- $v_1, \dots, v_m$ is linearly independent if and only if each vector in $\operatorname{span}(v_1, \dots, v_m)$ has only one representation as a linear combination of $v_1, \dots, v_m$.

- If some vectors are removed from a linearly independent list, the remaining list is also linearly independent.
  - Suppose $v_1, \dots, v_m$ is linearly independent. Suppose $v_1, \dots, v_n$ is not linearly independent, with $n < m$. Then $a_1 v_1 + \cdots + a_n v_n = 0$ for some $a_1, \dots, a_n \in \mathbb{F}$ such that $a_i \neq 0$ for all $i \in [n]$. But then $a_1 v_1 + \cdots + a_n v_n + 0 v_{n+1} + \cdots + 0 v_m = 0$, a contradiction.

- **Linearly dependent** (list $v_1, \dots, v_m$): A list $v_1, \dots, v_m$ of vectors in V that is not linearly independent.
  - In other words, $v_1, \dots, v_m$ are linearly dependent if there exist $a_1, \dots, a_m \in \mathbb{F}$, not all 0, such that $a_1 v_1 + \cdots + a_m v_m = 0$.

- The following is an important and oft-used lemma.

  Lemma 2.2 (Linear Dependence Lemma)**.** *Suppose $v_1, \dots, v_m$ is a linearly dependent list in V. Then there exists $j \in \{1, \dots, m\}$ such that the following hold:*

  (a) $v_j \in \operatorname{span}(v_1, \dots, v_{j-1})$*;*

  (b) if the j^{th} term is removed from $v_1, \dots, v_m$, the span of the remaining list equals $\operatorname{span}(v_1, \dots, v_m)$*.*

Proof. We divide into two cases (the list is $v_1 = 0$, and the list is $v_1, \dots, v_m$).

If the list is $v_1 = 0$, then the list is linearly dependent. Choose $j = 1$. Clearly, $v_1 \in \operatorname{span}() = \{0\}$ by definition. Additionally, $\operatorname{span}() = \{0\} = \{a_1 0 : a_1 \in \mathbb{F}\} = \operatorname{span}(v_1)$, as desired.

Since $v_1, \dots, v_m$ is linearly dependent, there exist $a_1, \dots, a_m \in \mathbb{F}$, not all 0, such that $a_1v_1 + \cdots + a_mv_m = 0$. Let j be the largest element of $\{1, \dots, m\}$ such that $a_j \neq 0$. Then

$$\begin{aligned} 0 &= a_1v_1 + \cdots + a_jv_j + 0v_{j+1} + \cdots + 0v_m \\ -a_jv_j &= a_1v_1 + \cdots + a_{j-1}v_{j-1} \\ v_j &= -\frac{a_1}{a_j}v_1 - \cdots - \frac{a_{j-1}}{a_j}v_{j-1} \end{aligned}$$

It follows that $v_j \in \operatorname{span}(v_1, \dots, v_{j-1})$, as desired.

Now clearly $\operatorname{span}(v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_m) \subset \operatorname{span}(v_1, \dots, v_m)$. In the other direction, suppose $u = c_1v_1 + \cdots + c_mv_m \in \operatorname{span}(v_1, \dots, v_m)$. Then

$$u = c_1v_1 + \cdots + c_{j-1}v_{j-1} + c_j\left(-\frac{a_1}{a_j}v_1 - \cdots - \frac{a_{j-1}}{a_j}v_{j-1}\right) + c_{j+1}v_{j+1} + \cdots + c_mv_m \in \operatorname{span}(v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_m)$$

as desired. ■

- We next prove an immediate consequence of the Linear Dependence Lemma.

  Theorem 2.3. *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 $u_1, \dots, u_m$ is linearly independent in V, and that $w_1, \dots, w_n$ spans V. We must prove that $m \leq n$. To do so, it will suffice to use the following m-step process.

  Step 1: Let $B = w_1, \dots, w_n$. Adding any $v \in V$ to this list produces a linearly dependent list (because the newly adjoined vector can be written as a linear combination of the other vectors). In particular, the list $u_1, w_1, \dots, w_n$ is linearly dependent. Thus, since $u_1 \neq 0$ (it's part of a linearly independent list, and thus cannot be written as $0u_i$ for any u_i), the Linear Dependence Lemma asserts that we can remove one of the w_i's such that the new list B consisting of u_1 and the remaining w_i's spans V.

  Step j: The list B from step $j-1$ spans V. Thus, as before, adjoin vector u_j to B, placing it just after $u_1, \dots, u_{j-1}$. It follows by the Linear Dependence Lemma that one of the following vectors (i.e., one of the w_i's) is in $\operatorname{span}(u_1, \dots, u_j)$, so we can remove it and know that the list comprised of $u_1, \dots, u_j$ followed by the remaining w_i's spans V.

  After step m, we have added all of the u's and the process stops. At each step, as we add a u to B, the Linear Dependence Lemma implies that there is some w to remove. Thus, there are at least as many w's as u's.[1] ■

  - This theorem allows us to prove results such as no list of 4 or more vectors is linearly independent in $\mathbb{R}^3$ (since $(1,0,0), (0,1,0), (0,0,1)$ spans $\mathbb{R}^3$), and no list of fewer than 4 vectors spans $\mathbb{R}^4$ (since $(1,0,0,0), (0,1,0,0), (0,0,1,0), (0,0,0,1)$ is linearly independent in $\mathbb{R}^4$).

- We now can rigorously prove that subspaces of finite-dimensional vector spaces are also finite-dimensional.

  Theorem 2.4. *Every subspace of a finite-dimensional vector space is finite-dimensional.*

  Proof. Let V be finite-dimensional, and suppose for the sake of contradiction that U is infinite-dimensional subspace of V. Since V is finite-dimensional, there exists a list of vectors $v_1, \dots, v_m$ such that $\operatorname{span}(v_1, \dots, v_m) = V$. To arrive at a contradiction, we will construct a linearly independent list of vectors in U of length $m + 1$, contradicting Theorem 2.3.

[1] We should be able to do this more rigorously via induction on m.

Since U is infinite-dimensional, there is no list of vectors in U spans it. Thus, if we choose $u_1 \in U$, we know that $\text{span}(u_1) \neq U$. It follows since $\text{span}(u_1) \subset U$ (as we know from the closure of U) that there exists $u_2 \in U$ such that $u_2 \notin \text{span}(u_1)$. However, we will still have that $\text{span}(u_1, u_2) \neq U$. More importantly, though, since $u_2 \notin \text{span}(u_1)$ and $u_1 \notin \text{span}()$, the Linear Dependence Lemma implies that u_1, u_2 is linearly independent. We can clearly continue in this fashion up to $u_1, \dots, u_{m+1}$, as desired. ■

2.B Bases

- **Basis** (of V): A list of vectors in V that is linearly independent and spans V.
- **Standard basis** (of $\mathbb{F}^n$): The list $(1, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, \dots, 0, 1)$.
- Determining whether a list of vectors is a basis:

Theorem 2.5. *A list $v_1, \dots, v_n$ of vectors in V is a basis of V if and only if every $v \in V$ can be written uniquely in the form*

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

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

Proof. Suppose first that $v_1, \dots, v_n$ is a basis of V. Let $v \in V$ be arbitrary. We will first show that v can be written in the desired form, and then we will show that this writing is unique. Let's begin. By the definition of a basis, $v_1, \dots, v_n$ spans V. Thus, $\text{span}(v_1, \dots, v_n) = V$. It follows that $v \in \text{span}(v_1, \dots, v_n)$, which implies by the definition of span that $v = a_1 v_1 + \cdots + a_n v_n$ where $a_1, \dots, a_n \in \mathbb{F}$, as desired. Now suppose for the sake of contradiction $v = c_1 v_1 + \cdots + c_n v_n$ as well, where $c_1, \dots, c_n \in \mathbb{F}$ and $c_j \neq a_j$ for some $i \in [n]$. Then

$$\begin{aligned} 0 &= v - v \\ &= (a_1 v_1 + \cdots + a_n v_n) - (c_1 v_1 + \cdots + c_n v_n) \\ &= (a_1 - c_1) v_1 + \cdots + (a_n - c_n) v_n \end{aligned}$$

Since at least $a_j - c_j \neq 0$ but the above sum still does equal 0, we have that $v_1, \dots, v_n$ are not linearly independent, a contradiction.

Now suppose that every $v \in V$ can be written uniquely in the form $v = a_1 v_1 + \cdots + a_n v_n$. To prove that $v_1, \dots, v_n$ is a basis of V, it will suffice to show that $v_1, \dots, v_n$ spans V and is linearly independent. Let's start with the first claim. Clearly, $\text{span}(v_1, \dots, v_n) \subset V$, and since every $v \in V$ may be written as a linear combination of $v_1, \dots, v_n$, we know that every $v \in V$ is an element of $\text{span}(v_1, \dots, v_n)$, as desired. On the other hand, we know that $0 = 0v_1 + \cdots + 0v_n$ and 0 can only be written in this unique form. Thus, the only choice of $a_1, \dots, a_m \in \mathbb{F}$ that makes $a_1 v_1 + \cdots + a_n v_n = 0$ is $a_1 = \cdots = a_n = 0$, proving that $v_1, \dots, v_n$ is linearly independent. ■

- Finding the basis in a spanning list.

Theorem 2.6. *Every spanning list in a vector space can be reduced to a basis of the vector space.*

Proof. Let $v_1, \dots, v_n$ span V. We induct on n. For the base case $n = 0$, if () spans V, then since () is linearly independent by definition, we are done. Now suppose inductively that we have proven that every spanning list in V of length n can be reduced to a basis of V; we wish to prove that every spanning list in V of length $n + 1$ can be reduced to a basis of V. Let $v_1, \dots, v_{n+1}$ span V. If $v_1, \dots, v_{n+1}$ is linearly independent, we are done. If $v_1, \dots, v_{n+1}$ is linearly dependent, then by the Linear Dependence Lemma, we can remove a vector from the list without changing its span. Our new list only has length n, so by the inductive hypothesis, it will reduce to a basis of V. ■

- Proving the existence of a basis in a finite-dimensional vector space.

Theorem 2.7. *Every finite-dimensional vector space has a basis.*

Proof. Let V be finite-dimensional. As such, there exists a list $v_1, \dots, v_n$ of vectors in V that spans V. It follows by Theorem 2.6 that some sublist of $v_1, \dots, v_n$ is a basis of V, as desired. ■

- Extending a linearly independent list into a basis (basically the dual of Theorem 2.6).

Theorem 2.8. *Every linearly independent list of vectors in a finite-dimensional vector space can be extended to a basis of the vector space.*

Proof. Let $u_1, \dots, u_m$ be a linearly independent list of vectors in V. By Theorem 2.7, V has a basis $w_1, \dots, w_n$. It follows that $u_1, \dots, u_m, w_1, \dots, w_n$ spans V. Thus, by Theorem 2.6, which removes the first linearly dependent vector in $u_1, \dots, u_m, w_1, \dots, w_n$ (necessarily one of the w_i's since $u_1, \dots, u_m$ are linearly independent) via the Linear Dependence Lemma, there exists a sublist of $u_1, \dots, u_m, w_1, \dots, w_n$ containing $u_1, \dots, u_m$ that is a basis of V. ■

- Finding orthogonal complements.

Theorem 2.9. *Suppose V is finite-dimensional and U is a subspace of V. Then there is a subspace W of V such that $V = U \oplus W$.*

Proof. Since V is finite-dimensional, Theorem 2.4 asserts that U is finite-dimensional. Thus, by Theorem 2.7, U has a basis $u_1, \dots, u_m$. It follows by Theorem 2.8 that there exist $w_1, \dots, w_n \in V$ such that $u_1, \dots, u_m, w_1, \dots, w_n$ is a basis of V. Let $W = \text{span}(w_1, \dots, w_n)$.

To prove that $U \oplus W = V$, it will suffice to show that

$$U + W = V \qquad\qquad U \cap W = \{0\}$$

To prove the first equation, it will suffice to show that for any vector $v \in V$, $v = u + w$ for $u \in U$ and $w \in W$. But since $u_1, \dots, u_m, w_1, \dots, w_n$ is a basis of V, we have by the definition of span that

$$v = \underbrace{a_1 u_1 + \cdots + a_m u_m}_{u} + \underbrace{b_1 w_1 + \cdots + b_n w_n}_{w}$$

as desired.

To prove the second equation, let $v \in U \cap W$ be arbitrary. Then since $v \in U$ and $u_1, \dots, u_m$ is a basis of U, we have that $v = a_1 u_1 + \cdots + a_m u_m$ where $a_1, \dots, a_m \in \mathbb{F}$. Similarly, we have that $v = b_1 w_1 + \cdots + b_m w_n$ where $b_1, \dots, b_n \in \mathbb{F}$. It follows that

$$\begin{aligned} 0 &= v - v \\ &= a_1 u_1 + \cdots + a_m u_m - b_1 w_1 - \cdots - b_n w_n \end{aligned}$$

But since $u_1, \dots, u_m, w_1, \dots, w_n$ is a basis of V, $u_1, \dots, u_m, w_1, \dots, w_n$ is linearly independent. It follows that $a_1 = \cdots = a_m = b_1 = \cdots = b_n = 0$. Therefore, $v = a_1 u_1 + \cdots + a_m u_m = 0$, as desired. ■

  - Note that this same basic idea extends to the infinite-dimensional case, but that proof requires considerably more advanced tools.

2.C Dimension

- It would be nice to define the dimension of a vector space as the length of a basis, but in order to do this we must first show that all bases have the same length.

Theorem 2.10. *Any two bases of a finite-dimensional vector space have the same length.*

Proof. Let B_1, B_2 be two arbitrary bases of V. Since B_1 is linearly independent in V and B_2 spans V, Theorem 2.3 asserts that $\operatorname{len} B_1 \leq \operatorname{len} B_2$. Similarly, since B_2 is linearly independent in V and B_1 spans V, Theorem 2.3 asserts that $\operatorname{len} B_2 \leq \operatorname{len} B_1$. Therefore, $\operatorname{len} B_1 = \operatorname{len} B_2$, as desired. ∎

- **Dimension** (of V finite-dimensional): The length of any basis of V. *Denoted by* $\mathbf{dim}\, \boldsymbol{V}$.
- We can now give the expected inequality regarding the dimension of subspaces with respect to the dimension of the vector space.

Theorem 2.11. *If V is finite-dimensional, and U is a subspace of V, then* $\dim U \leq \dim V$.

Proof. Since V is finite dimensional, Theorem 2.4 asserts that U is finite-dimensional. Thus, Theorem 2.7 implies that they have bases $B_U = u_1, \dots, u_m$ and $B_V = v_1, \dots, v_n$. Therefore, since B_U is linearly independent in V and B_V spans V, Theorem 2.3 asserts that $\dim U = \operatorname{len} B_U \leq \operatorname{len} B_V = \dim V$, as desired. ∎

- Note that while there exists an order-preserving bijection over addition and scalar multiplication by elements of $\mathbb{R}$ between $\mathbb{R}^2$ and $\mathbb{C}$, $\dim \mathbb{R}^2 = 2 \neq 1 = \dim \mathbb{C}$. Thus, when we talk about the dimension of a vector space, the role played by the choice of $\mathbb{F}$ cannot be neglected.
- Consider a list of vectors in a vector space that has the same length as a basis of the vector space. To check that this list is a basis, we can show that it suffices to show that the list is linearly independent *or* spans the vector space, as we will now prove.

Theorem 2.12. *Suppose V is finite-dimensional. Then every linearly independent list of vectors in V with length* $\dim V$ *is a basis of V.*

Proof. Let $\dim V = n$, and let $v_1, \dots, v_n$ be linearly independent. By Theorem 2.8, we can extend $v_1, \dots, v_n$ to a basis of V. However, since every basis of V has length n by Theorem 2.10, we need not add any vectors to $v_1, \dots, v_n$ to make it a basis; in other words, $v_1, \dots, v_n$ already is a basis. ∎

Theorem 2.13. *Suppose V is finite-dimensional. Then every spanning list of vectors in V with length* $\dim V$ *is a basis of V.*

Proof. The proof is symmetric to the proof of Theorem 2.12. ∎

- Lastly, we tackle the dimension of a sum of subspaces, which bears a resemblance to the principle of inclusion-exclusion.

Theorem 2.14. *If U_1 and U_2 are subspaces of a finite-dimensional vector space, then*

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2)$$

Proof. By Theorem 2.7, $U_1 \cap U_2$ (which we can prove is a subspace in its own right) has a basis, which we may denote $u_1, \dots, u_m$. Since $u_1, \dots, u_m$ is linearly independent in U_1, Theorem 2.8 asserts that it can be extended to a basis $u_1, \dots, u_m, v_1, \dots, v_j$ of U_1. Similarly, it can be extended to a basis $u_1, \dots, u_m, w_1, \dots, w_k$ of U_2.

To prove that $u_1, \dots, u_m, v_1, \dots, v_j, w_1, \dots, w_k$ is a basis of $U_1 + U_2$, it will suffice to show that it is linearly independent and spans $U_1 + U_2$.

To show that $u_1, \dots, u_m, v_1, \dots, v_j, w_1, \dots, w_k$ is linearly independent, it will suffice to verify that

$$\begin{aligned} a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_jv_j + c_1w_1 + \cdots + c_kw_k &= 0 \\ c_1w_1 + \cdots + c_kw_k &= -a_1u_1 - \cdots - a_mu_m - b_1v_1 - \cdots - b_jv_j \end{aligned}$$

implies that all of the coefficients equal 0. Suppose that the above equation holds. Then since $c_1w_1 + \cdots + c_kw_k$ can be written as a linear combination of the basis vectors of U_1, $c_1w_1 + \cdots + c_kw_k \in U_1$.

Additionally, since $c_1w_1 + \cdots + c_kw_k$ is a linear combination of vectors in U_2, $c_1w_1 + \cdots + c_kw_k \in U_2$. Thus, $c_1w_1 + \cdots + c_kw_k \in U_1 \cap U_2$. It follows that $c_1w_1 + \cdots + c_kw_k$ can be written as a linear combination of $u_1, \dots, u_m$, i.e.,

$$\begin{aligned} c_1w_1 + \cdots + c_kw_k &= d_1u_1 + \cdots + d_mu_m \\ 0 &= d_1u_1 + \cdots + d_mu_m - c_1w_1 - \cdots - c_kw_k \end{aligned}$$

for some $d_1, \dots, d_m \in \mathbb{F}$. But since $u_1, \dots, u_m, w_1, \dots, w_k$ is linearly independent as the basis of U_2, the above equation implies that $c_1 = \cdots = c_k = 0$. This implies that

$$0 = -a_1u_1 - \cdots - a_mu_m - b_1v_1 - \cdots - b_jv_j$$

meaning since $u_1, \dots, u_m, v_1, \dots, v_j$ is linearly independent as the basis of U_1, the above equation implies that $a_1 = \cdots = a_m = b_1 = \cdots = b_j = 0$, as desired.

To show that $u_1, \dots, u_m, v_1, \dots, v_j, w_1, \dots, w_k$ spans $U_1 + U_2$, it will suffice to show that all vectors in the list are elements of $U_1 + U_2$ (i.e., $\operatorname{span}(u_1, \dots, u_m, v_1, \dots, v_j, w_1, \dots, w_k) \subset U_1 + U_2$), and that every vector in $U_1 + U_2$ can be written as a linear combination of $u_1, \dots, u_m, v_1, \dots, v_j, w_1, \dots, w_k$ (i.e., that $U_1 + U_2 \subset \operatorname{span}(u_1, \dots, u_m, v_1, \dots, v_j, w_1, \dots, w_k)$). Since every vector in the list is an element of U_1 or U_2, we can show that it is an element of $U_1 + U_2$ by adding it to the additive identity of the other space. On the other hand, let $x \in U_1 + U_2$. Then $x = x_1 + x_2$, where $x_1 \in U_1$ and $x_2 \in U_2$. It follows that $x_1 = a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_jv_j$ and $x_2 = a_1'u_1 + \cdots + a_m'u_m + c_1w_1 + \cdots + c_kw_k$. Therefore, $x = (a_1 + a_1')u_1 + \cdots + (a_m + a_m')u_m + b_1v_1 + \cdots + b_jv_j + c_1w_1 + \cdots + c_kw_k$, as desired.

Having established that $u_1, \dots, u_m, v_1, \dots, v_j, w_1, \dots, w_k$ is a basis of $U_1 + U_2$, we have that

$$\begin{aligned} \dim(U_1 + U_2) &= m + j + k \\ &= (m + j) + (m + k) - m \\ &= \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2) \end{aligned}$$

as desired. ■

Exercises

1 Suppose V is finite-dimensional and U is a subspace of V such that $\dim U = \dim V$. Prove that $U = V$.

Proof. Since V is finite dimensional and U is a subspace of V, we have by Theorem 2.9 that there is a subspace W of V such that $V = U \oplus W$. It follows by Theorem 2.14 that

$$\dim W = \dim V - \dim U + \dim\{0\} = 0$$

Thus, $W = \{0\}$. Therefore,

$$\begin{aligned} V &= U \oplus W \\ &= \{u + w : u \in U, w \in W\} \\ &= \{u + 0 : u \in U\} \\ &= U \end{aligned}$$

as desired. ■

Chapter 3

Linear Maps

3.A The Vector Space of Linear Maps

9/5:

- **Linear map** (from V to W): A function $T : V \to W$ with the following properties. *Also known as* **linear transformation**.

  additivity
  $T(u + v) = Tu + Tv$ for all $u, v \in V$.

  homogeneity
  $T(\lambda v) = \lambda(Tv)$ for all $\lambda \in \mathbb{F}$ and all $v \in V$.

  - Note that for linear maps, Tv means the same as the more standard functional notation $T(v)$.
- $\boldsymbol{\mathcal{L}(V, W)}$: The set of all linear maps from V to W.
- **Zero map**: The function $0 \in \mathcal{L}(V, W)$ that takes each element of some vector space to the additive identity of another vector space. *Defined by*

$$0v = 0$$

- **Identity map**: The function $I \in \mathcal{L}(V, V)$ on some vector space that takes each element to itself. *Defined by*

$$Iv = v$$

- We can also formalize more esoteric linear processes as linear maps.
  - For example, $D \in \mathcal{L}(\mathcal{P}(\mathbb{R}), \mathcal{P}(\mathbb{R}))$ can be thought of as the differentiation map $Dp = p'$. This formalizes the fact that $(f + g)' = f' + g'$ and $(\lambda f)' = \lambda f'$.
  - We can do the same with integration: Let $T \in \mathcal{L}(\mathcal{P}(\mathbb{R}), \mathbb{R})$ be described by $Tp = \int_0^1 p(x)\,\mathrm{d}x$. This formalizes the fact that integrals are additive and homogenous.
  - Axler (2015) gives a number more examples.
- We now prove that there exists a unique linear map from a vector space of dimension n to any n vectors in another vector space.

  Theorem 3.1. *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 $Tv_j = w_j$ for each $j = 1, \dots, n$.*

  Proof. First, we define a function $T : V \to W$. We then show that T satisfies the specified property. After that, we show that it is a linear map. Lastly, we show that it is unique. Let's begin.

Let $T : V \to W$ be defined by

$$T(c_1v_1 + \cdots + c_nv_n) = c_1w_1 + \cdots + c_nw_n$$

for all $c_1v_1 + \cdots + c_nv_n \in V$. Note that this definition is valid since, by Theorem 2.5, each $v \in V$ can be written in the form $c_1v_1 + \cdots + c_nv_n$ where $c_1, \dots, c_n \in \mathbb{F}$.

To prove that $Tv_j = w_j$ for all $j = 1, \dots, n$, let each c_i in the above definition equal 0 save c_j, which we set equal to 1. Then we have

$$\begin{aligned} T(0v_1 + \cdots + 0v_{j-1} + 1v_j + 0v_{j+1} + \cdots + 0v_n) &= 0w_1 + \cdots + 0w_{j-1} + 1w_j + 0w_{j+1} + \cdots + 0w_n \\ T(v_j) &= w_j \end{aligned}$$

as desired.

To prove that T is a linear map, it will suffice to verify additivity and homogeneity, which we may do as follows. Let $u, v \in V$ with $u = a_1v_1 + \cdots + a_nv_n$ and $v = c_1v_1 + \cdots + c_nv_n$, and let $\lambda \in \mathbb{F}$. Then

$$\begin{aligned} T(u + v) &= T((a_1 + c_1)v_1 + \cdots + (a_n + c_n)v_n) \\ &= (a_1 + c_1)w_1 + \cdots + (a_n + c_n)w_n \\ &= Tu + Tv \end{aligned}$$

and

$$\begin{aligned} T(\lambda v) &= T(\lambda c_1v_1 + \cdots + \lambda c_nv_n) \\ &= \lambda c_1w_1 + \cdots + \lambda c_nw_n \\ &= \lambda Tv \end{aligned}$$

as desired.

Now suppose $\tilde{T} \in \mathcal{L}(V, W)$ satisfies $\tilde{T}v_j = w_j$ for all $j = 1, \dots, n$. To prove that $T = \tilde{T}$, it will suffice to show that $\tilde{T}(c_1v_1 + \cdots + c_nv_n) = T(c_1v_1 + \cdots + c_nv_n)$ for all $c_1v_1 + \cdots + c_nv_n \in V$. Let $c_1v_1 + \cdots + c_nv_n \in V$ be arbitrary. We know that $\tilde{T}(v_j) = w_j$ for all $j = 1, \dots, n$. It follows since $\tilde{T}$ is a linear map (specifically, since it's homogenous) that $c_jw_j = c_j\tilde{T}(v_j) = \tilde{T}(c_jv_j)$ for all $j = 1, \dots, n$. Similarly, the additivity of $\tilde{T}$ implies that

$$\begin{aligned} T(c_1v_1 + \cdots + c_nv_n) &= c_1w_1 + \cdots + c_nw_n \\ &= \tilde{T}(c_1v_1) + \cdots + \tilde{T}(c_nv_n) \\ &= \tilde{T}(c_1v_1 + \cdots + c_nv_n) \end{aligned}$$

as desired. ■

- **Sum** (of $S, T \in \mathcal{L}(V, W)$): The linear map $(S + T) \in \mathcal{L}(V, W)$ defined by $(S + T)(v) = Sv + Tv$ for all $v \in V$.
- **Product** (of $T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbb{F}$): The linear map $(\lambda T) \in \mathcal{L}(V, W)$ defined by $(\lambda T)(v) = \lambda(Tv)$ for all $v \in V$.
- It follows that, under these definitions of addition and multiplication, $\mathcal{L}(V, W)$ is a vector space.
- **Product** (of $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$): The linear map $ST \in \mathcal{L}(U, W)$ defined by $(ST)(u) = S(Tu)$ for all $u \in U$.
  - Note that the product is just function composition, but most mathematicians do write ST instead of $S \circ T$.
- Linear maps (with the correct corresponding domains) satisfy the associativity, identity, and distributive properties. However, multiplication of linear maps is not necessarily commutative.

  - $(T_1T_2)T_3 = T_1(T_2T_3)$.
  - $TI_V = I_WT = T$ (note that if $T \in \mathcal{L}(V, W)$, $I_V \in \mathcal{L}(V, V)$ and $I_W \in \mathcal{L}(W, W)$).
  - $(S_1 + S_2)T = S_1T + S_2T$ and $S(T_1 + T_2) = ST_1 + ST_2$.

- Linear maps send 0 to 0.

  Theorem 3.2. *Suppose $T \in \mathcal{L}(V, W)$. Then $T(0) = 0$.*

  Proof. By additivity, we have

  $$\begin{aligned} T(0) &= T(0 + 0) = T(0) + T(0) \\ 0 &= T(0) \end{aligned}$$

  as desired. ∎

3.B Null Spaces and Ranges

- **Null space** (of $T \in \mathcal{L}(V, W)$): The subset of V consisting of those vectors that T maps to 0. *Also known as* **kernel**. *Denoted by* $\textbf{null}\,\boldsymbol{T}$. *Given by*

  $$\operatorname{null} T = \{v \in V : Tv = 0\}$$

- The null space is a subspace.

  Theorem 3.3. *Suppose $T \in \mathcal{L}(V, W)$. Then* $\operatorname{null} T$ *is a subspace of V.*

  Proof. To prove that $\operatorname{null} T$ is a subspace of V, it will suffice to show that $0 \in \operatorname{null} T$, $u, v \in \operatorname{null} T$ implies that $u + v \in \operatorname{null} T$, and $u \in \operatorname{null} T$ and $\lambda \in \mathbb{F}$ imply $\lambda u \in \operatorname{null} T$. Let's begin.

  By Theorem 3.2, $T(0) = 0$. Therefore, $0 \in \operatorname{null} T$, as desired.

  Let $u, v \in \operatorname{null} T$ be arbitrary. Then by additivity

  $$T(u + v) = Tu + Tv = 0 + 0 = 0$$

  so $u + v \in \operatorname{null} T$, as desired.

  Let $u \in \operatorname{null} T$ and $\lambda \in \mathbb{F}$ be arbitrary. Then by homogeneity,

  $$T(\lambda u) = \lambda Tu = \lambda 0 = 0$$

  so $\lambda u \in \operatorname{null} T$, as desired. ∎

- **Injective** (function): A function $T : V \to W$ such that $Tu = Tv$ implies $u = v$. *Also known as* **one-to-one**.

- If 0 is the only vector that gets mapped to 0, then T is injective.

  Theorem 3.4. *Let $T \in \mathcal{L}(V, W)$. Then T is injective if and only if* $\operatorname{null} T = \{0\}$.

  Proof. Suppose first that T is injective. To prove that $\operatorname{null} T = \{0\}$, it will suffice to show that $0 \in \operatorname{null} T$ and for every $v \in \operatorname{null} T$, $v = 0$. By Theorem 3.3, $0 \in \operatorname{null} T$. Now let $v \in \operatorname{null} T$ be arbitrary. By the definition of the null space, we have $Tv = 0$. By Theorem 3.2, we have $T(0) = 0$. Thus, by transitivity, we have that $Tv = T(0)$. It follows by injectivity that $v = 0$, as desired.

  Now suppose that $\operatorname{null} T = \{0\}$. To prove that T is injective, it will suffice to show that if $Tu = Tv$, then $u = v$. Suppose $u, v \in V$ satisfy $Tu = Tv$. Then

  $$0 = Tu - Tv = T(u - v)$$

  so $(u - v) \in \operatorname{null} T = \{0\}$. It follows that $u - v = 0$, i.e., that $u = v$, as desired. ∎

- **Range** (of $T \in \mathcal{L}(V,W)$): The subset of W consisting of those vectors that are of the form Tv for some $v \in V$. *Also known as* **image**. *Denoted by* **range *T***. *Given by*

$$\operatorname{range} T = \{Tv : v \in V\}$$

- The range is a subspace.

Theorem 3.5. *Suppose $T \in \mathcal{L}(V,W)$. Then* range T *is a subspace of W.*

Proof. To prove that range T is a subspace of W, it will suffice to show that $0 \in \operatorname{range} T$, $w_1, w_2 \in \operatorname{range} T$ implies that $(w_1 + w_2) \in \operatorname{range} T$, and $w \in \operatorname{range} T$ and $\lambda \in \mathbb{F}$ imply $\lambda w \in \operatorname{range} T$. Let's begin.

By the definition of a vector space, $0 \in V$. By Theorem 3.2, $T(0) = 0$. Therefore, $0 \in \operatorname{range} T$, as desired.

Let $w_1, w_2 \in \operatorname{range} T$ be arbitrary. Then there exist $v_1, v_2 \in V$ such that $Tv_1 = w_1$ and $Tv_2 = w_2$. It follows by additivity that

$$T(v_1 + v_2) = Tv_1 + Tv_2 = w_1 + w_2$$

Therefore, since $v_1 + v_2 \in V$, we have that $(w_1 + w_2) \in \operatorname{range} T$, as desired.

Let $w \in \operatorname{range} T$ and $\lambda \in \mathbb{F}$ be arbitrary. Then there exists $v \in V$ such that $Tv = w$. It follows by homogeneity that

$$T(\lambda v) = \lambda Tv = \lambda w$$

Therefore, since $\lambda v \in V$, we have that $\lambda w \in \operatorname{range} T$, as desired. ■

- **Surjective** (function): A function $T : V \to W$ such that $\operatorname{range} T = W$. *Also known as* **onto**.

- We now prove a very important theorem.

Theorem 3.6 (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. By Theorem 3.3, null T is a subspace of V finite-dimensional. Thus, by Theorem 2.4, null T is finite-dimensional. It follows by Theorem 2.7 that we may let $u_1, \dots, u_m$ be a basis of null T. As a basis of a subspace of V, $u_1, \dots, u_m$ is a linearly independent list of vectors in V. Consequently, by Theorem 2.8, we may extend it to a basis $u_1, \dots, u_m, v_1, \dots, v_n$ of V.

Having established this terminology, we can now see that to prove that range T is finite-dimensional, it will suffice to show that $Tv_1, \dots, Tv_n$ spans it. To show that $\operatorname{span}(Tv_1, \dots, Tv_n) = \operatorname{range} T$, it will suffice to show that every $b_1Tv_1 + \cdots + b_nTv_n \in \operatorname{span}(Tv_1, \dots, Tv_n)$ is an element of range T and that every $Tv \in \operatorname{range} T$ is an element of $\operatorname{span}(Tv_1, \dots, Tv_n)$. Let $b_1Tv_1 + \cdots + b_nTv_n \in \operatorname{span}(Tv_1, \dots, Tv_n)$ be arbitrary. Then

$$\begin{aligned}
b_1Tv_1 + \cdots + b_nTv_n &= T(b_1v_1 + \cdots + b_nv_n) \\
&= T(0u_1 + \cdots + 0u_m + b_1v_1 + \cdots + b_nv_n)
\end{aligned}$$

Therefore, since $0u_1 + \cdots + 0u_m + b_1v_1 + \cdots + b_nv_n \in V$ by V's closure under addition and scalar multiplication, we have that $b_1Tv_1 + \cdots + b_nTv_n \in \operatorname{range} T$, as desired. Now let $Tv \in \operatorname{range} T$ be arbitrary. Since $v \in V$ and $u_1, \dots, u_m, v_1, \dots, v_n$ is a basis of V, Theorem 2.5 implies that $v = a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_nv_n$ for some $a_1, \dots, a_m, b_1, \dots, b_n \in \mathbb{F}$. Therefore,

$$\begin{aligned}
Tv &= T(a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_nv_n) \\
&= T(a_1u_1 + \cdots + a_mu_m) + T(b_1v_1 + \cdots + b_nv_n) \\
&= a_1Tu_1 + \cdots + a_mTu_m + b_1Tv_1 + \cdots + v_nTv_n \\
&= a_10 + \cdots + a_m0 + b_1Tv_1 + \cdots + v_nTv_n \\
&= b_1Tv_1 + \cdots + v_nTv_n
\end{aligned}$$

where each $Tu_j = 0$ because each $u_j \in \operatorname{null} T$, so $Tv \in \operatorname{span}(Tv_1, \dots, Tv_n)$, as desired.

Before we can verify the equation from the theorem, we need to establish one last fact: that $Tv_1, \dots, Tv_n$ is linearly independent. Suppose $c_1, \dots, c_n \in \mathbb{F}$ make

$$\begin{aligned} c_1 Tv_1 + \cdots + c_n Tv_n &= 0 \\ T(c_1 v_1 + \cdots + c_n v_n) &= 0 \end{aligned}$$

It follows that $c_1 v_1 + \cdots + c_n v_n \in \operatorname{null} T$. Thus, since $u_1, \dots, u_m$ is a basis of $\operatorname{null} T$ by Theorem 2.5, we have that

$$\begin{aligned} c_1 v_1 + \cdots + c_n v_n &= d_1 u_1 + \cdots + d_m u_m \\ 0 &= d_1 u_1 + \cdots + d_m u_m - c_1 v_1 - \cdots - c_n v_n \end{aligned}$$

for some $d_1, \dots, d_m \in \mathbb{F}$. But since $u_1, \dots, u_m, v_1, \dots, v_n$ is linearly independent as the basis of V, the above equation implies that $c_1 = \cdots = c_n = 0$, as desired.

Having established that $u_1, \dots, u_m, v_1, \dots, v_n$ is a basis of V, $u_1, \dots, u_m$ is a basis of $\operatorname{null} T$, and $Tv_1, \dots, Tv_n$ spans $\operatorname{range} T$ and is linearly independent in $\operatorname{range} T$ (i.e., is a basis of $\operatorname{range} T$), we have that

$$\begin{aligned} \dim V &= m + n \\ &= \dim \operatorname{null} T + \dim \operatorname{range} T \end{aligned}$$

as desired. ■

- We can now prove that a linear map to a "smaller" vector space cannot be injective.

Theorem 3.7. *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. Let $T \in \mathcal{L}(V, W)$. Then

$$\begin{aligned} \dim \operatorname{null} T &= \dim V - \dim \operatorname{range} T && \text{Fundamental Theorem of Linear Maps} \\ &\geq \dim V - \dim \operatorname{range} T && \text{Theorem 2.11} \\ &> 0 \end{aligned}$$

It follows that $\operatorname{null} T$ has a basis consisting of a list of one or more vectors. As a linearly independent list, naturally none of these vectors will be equal to the zero vector. Thus, since $\operatorname{null} T$ contains vectors other than 0, Theorem 3.4 implies that T is not injective. ■

- Similarly, we can prove that a linear map to a "bigger" vector space cannot be surjective.

Theorem 3.8. *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.*

Proof. Let $T \in \mathcal{L}(V, W)$. Then

$$\begin{aligned} \dim \operatorname{range} T &= \dim V - \dim \operatorname{null} T && \text{Fundamental Theorem of Linear Maps} \\ &\leq \dim V && < \dim W \end{aligned}$$

Therefore, $\operatorname{range} T \neq W$, so T cannot be surjective. ■

- Theorems 3.7 and 3.8 allow us to express questions about systems of linear equations in terms of linear maps.

  - For example, a question about a **homogenous** system of linear equations could be, "does there exist a nonzero solution to the homogenous system $\sum_{k=1}^n A_{1,k}x_k = 0, \dots, \sum_{k=1}^n A_{m,k}x_k = 0$?"
  - If we define $T : \mathbb{F}^n \to \mathbb{F}^m$ by

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

    we can express the system of equations as $T(x_1, \dots, x_n) = 0$ and ask instead, "is $\dim \operatorname{null} T > 0$?"

- **Homogenous** (system of linear equations): A system of m linear equations $\sum_{k=1}^n A_{1,k}x_k = c_1$ through $\sum_{k=1}^n A_{m,k}x_k = c_m$ such that the constant term $c_j = 0$ for all $j = 1, \dots, m$.
- Continuing with the linear equations example, we can rigorously show the following.

  Theorem 3.9. *A homogenous system of linear equations with more variables than equations has nonzero solutions.*

  Proof. In terms of the above, $T : \mathbb{F}^n \to \mathbb{F}^m$ where $n > m$. Thus, by Theorem 3.7, T is not injective. Consequently, by Theorem 3.4, $\dim \operatorname{null} T > 0$. Therefore, the system has nonzero solutions. ∎

  Theorem 3.10. *An inhomogenous system of linear equations with more equations than variables has no solution for some choice of constant terms.*

  Proof. In terms of the above, $T : \mathbb{F}^n \to \mathbb{F}^m$ where $m > n$. We want to know if there exists $(c_1, \dots, c_m) \in \mathbb{F}^m$ such that $T(x_1, \dots, x_n) \neq (c_1, \dots, c_m)$ for any $(x_1, \dots, x_n) \in \mathbb{F}^n$. In other words, we want to know if there exists $(c_1, \dots, c_m) \in \mathbb{F}^m$ such that $(c_1, \dots, c_m) \notin \operatorname{range} T$, i.e., if $\operatorname{range} T \neq \mathbb{F}^m$. But since $n < m$, Theorem 3.8 asserts that T is not surjective, meaning that $\operatorname{range} T \neq W$, as desired. ∎

  - Note that while the past two results are typically proven with Gaussian elimination, the abstract approach taken here leads to cleaner proofs.

3.C Matrices

- ***m*-by-*n* matrix**: A rectangular array A of elements of $\mathbb{F}$ with m rows and n columns

$$A = \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{pmatrix}$$

  where m and n are positive integers.

  - The notation $A_{j,k}$ denotes the entry in row j, column k of A. In other words, the first index refers to the row nunber and the second index refers to the column number.

- **Matrix** (of $T \in \mathcal{L}(V, W)$ with respect to the bases $v_1, \dots, v_n$ of V and $w_1, \dots, w_m$ of W): 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 are not clear from context, then the notation $\mathcal{M}(T, (v_1, \dots, v_n), (w_1, \dots, w_m))$ is used.
  - Another way of wording the definition states that the k^{th} column of $\mathcal{M}(T)$ consists of the scalars needed to write Tv_k as a linear combination of $w_1, \dots, w_m$.
  - Assuming standard bases, we "can think of the k^{th} column of $\mathcal{M}(T)$ as the T applied to the k^{th} standard basis vector" (Axler, 2015, p. 71).

- **Sum** (of two m-by-n matrices A, C): The m-by-n matrix $A + C$ defined by $(A + C)_{j,k} = A_{j,k} + C_{j,k}$.
  - Symbolically,

$$\begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{pmatrix} + \begin{pmatrix} C_{1,1} & \cdots & C_{1,n} \\ \vdots & & \vdots \\ C_{m,1} & \cdots & C_{m,n} \end{pmatrix} = \begin{pmatrix} A_{1,1} + C_{1,1} & \cdots & A_{1,n} + C_{1,n} \\ \vdots & & \vdots \\ A_{m,1} + C_{m,1} & \cdots & A_{m,n} + C_{m,n} \end{pmatrix}$$

- Suppose $S, T \in \mathcal{L}(V, W)$. Then $\mathcal{M}(S + T) = \mathcal{M}(S) + \mathcal{M}(T)$.
- **Product** (of an m-by-n matrix A and $\lambda \in \mathbb{F}$): The m-by-n matrix λA defined by $(\lambda A)_{j,k} = \lambda A_{j,k}$.
  - Symbolically,

$$\lambda \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{pmatrix} = \begin{pmatrix} \lambda A_{1,1} & \cdots & \lambda A_{1,n} \\ \vdots & & \vdots \\ \lambda A_{m,1} & \cdots & \lambda A_{m,n} \end{pmatrix}$$

- Suppose $\lambda \in \mathbb{F}$ and $T \in \mathcal{L}(V, W)$. Then $\mathcal{M}(\lambda T) = \lambda \mathcal{M}(T)$.
- $\boldsymbol{\mathbb{F}^{m,n}}$: The set of all m-by-n matrices with entries in $\mathbb{F}$, where m and n are positive integers.
- We have that $\dim \mathbb{F}^{m,n} = mn$.
  - Note that a basis of $\mathbb{F}^{m,n}$ is the set of all m-by-n matrices that have 0s everywhere save a 1 in a single place.
- **Product** (of an m-by-n matrix A and an n-by-p matrix C): The m-by-p matrix AC defined by $(AC)_{j,k} = \sum_{r=1}^{n} A_{j,r} C_{r,k}$.
  - We may derive this by noting that if $\mathcal{M}(S) = A$ and $\mathcal{M}(T) = C$, $T : U \to V$ and $S : V \to W$, and $u_1, \dots, u_p$, $v_1, \dots, v_n$, and $w_1, \dots, w_m$ are bases, then

$$\begin{aligned} (ST)u_k &= S\left(\sum_{r=1}^{n} C_{r,k} v_r\right) \\ &= \sum_{r=1}^{n} C_{r,k} S v_r \\ &= \sum_{r=1}^{n} C_{r,k} \sum_{j=1}^{m} A_{j,r} w_j \\ &= \sum_{j=1}^{m} \left(\sum_{r=1}^{n} A_{j,r} C_{r,k}\right) w_j \end{aligned}$$

  - Matrix multiplication is not commutative, but is distributive and associative.
- We now prove a relation between matrix and operator composition.

  Theorem 3.11. *Suppose $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$. Then $\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T)$.*

- If A is an m-by-n matrix, then...
  - We let $A_{j,\cdot}$ denote the 1-by-n matrix consisting of row j of A;
  - We let $A_{\cdot,k}$ denote the m-by-1 matrix consisting of column k of A.
- Thus, if A is an m-by-n matrix and C is an n-by-p matrix, then $(AC)_{j,k} = A_{j,\cdot} C_{\cdot,k}$ for all $1 \leq j \leq m$ and $1 \leq k \leq p$.

- Similarly, $(AC)_{\cdot,k} = AC_{\cdot,k}$.
- Lastly, suppose A is an m-by-n matrix and $c = (c_1, \dots, c_n)$ is an n-by-1 matrix. Then $Ac = c_1 A_{\cdot,1} + \cdots + c_n A_{\cdot,n}$.
  - In other words, Ac is a linear combination of the columns of A, with the scalars that multiply the columns coming from c.

3.D Invertibility and Isomorphic Vector Spaces

9/6:

- **Invertible** (linear map): A linear map $T \in \mathcal{L}(V, W)$ such that there exists a linear map $S \in \mathcal{L}(V, W)$ such that ST equals the identity map on V and TS equals the identity map on W.
- **Inverse** (of $T \in \mathcal{L}(V, W)$): The linear map $S \in \mathcal{L}(W, V)$ satisfying $ST = I_V$ and $TS = I_W$. *Denoted by* $\boldsymbol{T^{-1}}$.
- We now justify the use of the word "the" in the definition of the inverse.

  Theorem 3.12. *An invertible linear map has a unique inverse.*

  Proof. Suppose $T \in \mathcal{L}(V, W)$ is invertible and S_1, S_2 are inverses of T. Then

  $$S_1 = S_1 I = S_1(TS_2) = (S_1 T)S_2 = IS_2 = S_2$$

  as desired. ■

- We now give a criterion for invertibility.

  Theorem 3.13. *A linear map is invertible if and only if it is injective and surjective.*

  Proof. Let $T \in \mathcal{L}(V, W)$. Suppose first that T is invertible.

  To prove that T is injective, it will suffice to show that for all $u, v \in V$, $Tu = Tv$ implies that $u = v$. Let u, v be arbitrary elements of V that satisfy $Tu = Tv$. Then

  $$u = T^{-1}(Tu) = T^{-1}(Tv) = v$$

  as desired.

  To prove that T is surjective, it will suffice to show that $\operatorname{range} T = W$. Since $\operatorname{range} T \subset W$, we need only show that $W \subset \operatorname{range} T$. Let $w \in W$ be arbitrary. Since $w = T(T^{-1}w)$ where $T^{-1}w \in V$, we have that $w \in \operatorname{range} T$, as desired.

  Now suppose that T is injective and surjective. To prove that T is invertible, we will define a function $S : W \to V$, prove that it is a linear map, prove that $TS = I_W$, and prove that $ST = I_V$. Let Sw be the unique element of V such that $T(Sw) = w$ (the surjectivity of T guarantees that there exists *an* element of V that T maps to w, and the injectivity of T guarantees the uniqueness of said element).

  To prove that S is a linear map, it will suffice to show that S is additive and homogenous. To verify additivity, first note that the additivity of T implies that

  $$T(Sw_1 + Sw_2) = T(Sw_1) + T(Sw_2) = w_1 + w_2$$

  But since the above equation implies that $Sw_1 + Sw_2$ is the unique element of V that T maps to $w_1 + w_2$, we have by the definition of S that $S(w_1 + w_2) = Sw_1 + Sw_2$. The proof is symmetric for homogeneity.

  To prove that $TS = I_W$, we need only appeal to the definition of S, which states that $(TS)w = T(Sw) = w$ for all $w \in W$. It immediately follows that $TS = I_W$.

  To prove that $ST = I_V$, first note that for all $v \in V$,

  $$T((ST)v) = (TS)(Tv) = I(Tv) = Tv$$

  It follows by the injectivity of T that $(ST)v = v$, i.e., that $ST = I_V$, as desired. ■

- **Isomorphism**: An invertible linear map.
- **Isomorphic** (vector spaces): Two vector spaces such that there exists an isomorphism from one vector space onto the other one.
- Isomorphic vector spaces have the same dimension.

Theorem 3.14. *Two finite-dimensional vector spaces over $\mathbb{F}$ are isomorphic if and only if they have the same dimension.*

Proof. Suppose V, W are isomorphic finite-dimensional vector spaces over $\mathbb{F}$. Then there exists an isomorphism $T : V \to W$. By the definition of isomorphism, T is an invertible linear map, meaning by Theorem 3.13 that T is injective and surjective. Thus, since there exists an injective linear map $T : V \to W$, the contrapositive of Theorem 3.7 asserts that $\dim V \leq \dim W$. Additionally, since there exists a surjective linear map $T : V \to W$, the contrapositive of Theorem 3.8 asserts that $\dim V \geq \dim W$. Therefore, we have that $\dim V = \dim W$, as desired.

Now suppose that $\dim V = \dim W$. Let $v_1, \dots, v_n$ be a basis of V, and let $w_1, \dots, w_n$ be a basis of W. By Theorem 3.1, there exists a unique linear map $T : V \to W$ such that $Tv_j = w_j$ for each $j = 1, \dots, n$. To prove that T is an isomorphism, Theorem 3.13 tells us that it will suffice to show that it is injective and surjective. To show that T is surjective, it will suffice to show that $\operatorname{range} T = W = \operatorname{span}(w_1, \dots, w_n)$. But since $Tv_j = w_j \in \operatorname{range} T$ for all $j = 1, \dots, n$, $\operatorname{range} T \subset W$, and $\operatorname{range} T$ is a vector space (see Theorem 3.5), we have that $\operatorname{range} T = \operatorname{span}(w_1, \dots, w_n) = W$, as desired. To prove that T is injective, Theorem 3.4 tells us that it will suffice to show that $\operatorname{null} T = \{0\}$, i.e., that $\dim \operatorname{null} T = 0$. But since $\dim \operatorname{range} T = \dim W = \dim V$, we have by the Fundamental Theorem of Linear Maps that

$$\begin{aligned}
\dim \operatorname{null} T + \dim \operatorname{range} T &= \dim V \\
&= \dim W \\
&= \dim \operatorname{range} T \\
\dim \operatorname{null} T &= 0
\end{aligned}$$

as desired. ■

- This result implies that every finite-dimensional vector space of dimension n is isomorphic to $\mathbb{F}^n$.
- It also allows us to formalize the link between linear maps from V to W and matrices in $\mathbb{F}^{m,n}$.

Theorem 3.15. *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. We have already established that $\mathcal{M}(S + T) = \mathcal{M}(S) + \mathcal{M}(T)$ and that $\mathcal{M}(\lambda T) = \lambda \mathcal{M}(T)$, so we already know that $\mathcal{M}$ is a linear map. To prove that it is invertible, Theorem 3.13 tells us that it will suffice to show that $\mathcal{M}$ is injective and surjective.

To show that $\mathcal{M}$ is injective, Theorem 3.4 tells us that it will suffice to verify that $\operatorname{null} \mathcal{M} = \{0\}$. Let $T \in \mathcal{L}(V, W)$ be arbitrary. If $\mathcal{M}(T) = 0$ (0 denoting the zero matrix), then

$$Tv_k = 0w_1 + \cdots + 0w_m = 0$$

for all $k = 1, \dots, n$. But since $v_1, \dots, b_n$ is a basis of V, this implies that $T = 0$ (0 denoting the zero transformation), as desired.

To show that $\mathcal{M}$ is surjective, it will suffice to verify that $\operatorname{range} \mathcal{M} = \mathbb{F}^{m,n}$. Clearly $\operatorname{range} \mathcal{M} \subset \mathbb{F}^{m,n}$, so we focus on the other direction. Let $A \in \mathbb{F}^{m,n}$ be arbitrary. Define $T \in \mathcal{L}(V, W)$ by

$$Tv_k = \sum_{j=1}^{m} A_{j,k} w_j$$

for $k = 1, \dots, n$. It follows by the definition of a matrix of a linear transformation that $\mathcal{M}(T) = A$, as desired. ■

- We can now determine the dimension of the vector space of linear maps from one finite-dimensional vector space to another.

  Theorem 3.16. *Suppose V and W are finite-dimensional. Then $\mathcal{L}(V,W)$ is finite-dimensional and*

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

  Proof. By Theorem 3.15, $\mathcal{L}(V,W)$ and $\mathbb{F}^{m,n}$ are isomorphic. Thus, by Theorem 3.14, $\mathcal{L}(V,W)$ and $\mathbb{F}^{m,n}$ have the same dimension. Therefore, we have that

  $$\begin{aligned}\dim \mathcal{L}(V,W) &= \dim \mathbb{F}^{m,n}\\ &= mn\\ &= (\dim V)(\dim W)\end{aligned}$$

  as desired. ■

- **Matrix** (of $v \in V$ with respect to the basis $v_1,\dots,v_n$ of V): The n-by-1 matrix $\mathcal{M}(v)$ whose entries $A_{j,1}$ are defined by

  $$v = A_{1,1}v_1 + \cdots + A_{n,1}v_n$$

- We now show that the columns of the matrix of T are directly related to the effect T has on basis vectors.

  Theorem 3.17. *Suppose $T \in \mathcal{L}(V,W)$, $v_1,\dots,v_n$ is a basis of V, and $w_1,\dots,w_m$ is a basis of W. Let $1 \leq k \leq n$. Then*

  $$\mathcal{M}(T)_{\cdot,k} = \mathcal{M}(Tv_k)$$

  Proof. As an element of W, $Tv_k = c_1w_1 + \cdots + c_mw_m$ for some $c_1,\dots,c_m \in \mathbb{F}$. By the definition of the matrix of T, the values in column k are $c_1,\dots,c_m$. Similarly, by the definition of the matrix of Tv_k, the values in its one column are $c_1,\dots,c_m$, as desired. ■

- Linear maps act like matrix multiplication.

  Theorem 3.18. *Suppose $T \in \mathcal{L}(V,W)$, $v \in V$, $v_1,\dots,v_n$ is a basis of V, and $w_1,\dots,w_m$ is a basis of W. Then*

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

  Proof. Let $v = c_1v_1 + \cdots + c_nv_n$. Then by the linearity of T, $Tv = c_1Tv_1 + \cdots + c_nTv_n$. It follows by the linearity of $\mathcal{M}$, Theorem 3.17, and the fact that a matrix times a vector is equal to a linear combination of columns that

  $$\begin{aligned}\mathcal{M}(Tv) &= c_1\mathcal{M}(Tv_1) + \cdots + c_n\mathcal{M}(Tv_n)\\ &= c_1\mathcal{M}(T)_{\cdot,1} + \cdots + c_n\mathcal{M}(T)_{\cdot,n}\\ &= \mathcal{M}(T)\mathcal{M}(v)\end{aligned}$$

  as desired. ■

- "Each m-by-n matrix A induces a linear map from $\mathbb{F}^{n,1}$ to $\mathbb{F}^{m,1}$, namely the matrix multiplication function that takes $x \in \mathbb{F}^{n,1}$ to $Ax \in \mathbb{F}^{m,1}$" (Axler, 2015, p. 85).
- **Operator**: A linear map from a vector space to itself.
- $\boldsymbol{\mathcal{L}(V)}$: The set of all operators on V.
  - Mathematically, $\mathcal{L}(V) = \mathcal{L}(V,V)$.

- On infinite-dimensional vector spaces, neither injectivity nor surjectivity, alone, implies invertibility.
- However, only one does on finite-dimensional vector spaces.

Theorem 3.19. *Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Then the following are equivalent.*

(a) *T is invertible;*
(b) *T is injective;*
(c) *T is surjective.*

Proof. First, suppose that T is invertible. Then by Theorem 3.13, T is injective, as desired.

Second, suppose that T is injective. Then by Theorem 3.4, $\operatorname{null} T = \{0\}$. It follows by the Fundamental Theorem of Linear Maps that

$$\begin{aligned}\dim \operatorname{range} T &= \dim V - \dim \operatorname{null} T \\ &= \dim V\end{aligned}$$

Thus, since $\operatorname{range} T$ has the same dimension as V and is a subspace of V (by Theorem 3.5), $\operatorname{range} T = V$. Therefore, T is surjective, as desired.

Third, suppose that T is surjective. Then $\operatorname{range} T = V$. It follows that $\dim \operatorname{range} T = \dim V$. Thus, by the Fundamental Theorem of Linear Maps, we have that

$$\begin{aligned}\dim \operatorname{null} T &= \dim V - \dim \operatorname{range} T \\ &= 0\end{aligned}$$

Consequently, by Theorem 3.4, T is injective. Therefore, by Theorem 3.13, T is invertible, as desired. ∎

Exercises

10/11: **9** Suppose V is finite-dimensional and $S, T \in \mathcal{L}(V)$. Prove that ST is invertible if and only if both S and T are invertible.

Proof. Suppose first that ST is invertible. Then by Theorem 3.19, ST is injective and surjective. We will prove that S, T are invertible in turn; in particular, we will first prove that S is surjective and then prove that T is injective. Let's begin. To prove that S is surjective, it will suffice to show that $\operatorname{range} S = V$. Let $v \in V$ be arbitrary. Since ST is surjective, there exists $v' \in V$ such that $STv' = v$. Thus, since $Tv' \in V$, the fact that $S(Tv') = v$ implies that $v \in \operatorname{range} S$. The inclusion in the other direction is obvious. Now to prove that T is injective, it will suffice to show that $Tv = Tv'$ implies $v' = v'$. Let $Tv = Tv'$. Then $STv = STv'$. It follows that $v = v'$ by the injectivity of ST, as desired.

Now suppose that S and T are invertible. Then by Theorem 3.12, there exist S^{-1}, T^{-1} such that

$$SS^{-1} = I = S^{-1}S \qquad TT^{-1} = I = T^{-1}T$$

Let $(ST)^{-1} = T^{-1}S^{-1}$. Then

$$ST(ST)^{-1} = STT^{-1}S^{-1} = SIS^{-1} = SS^{-1} = I = T^{-1}T = T^{-1}IT = T^{-1}S^{-1}ST = (ST)^{-1}ST$$

so ST is invertible, as desired. ∎

10 Suppose V is finite-dimensional and $S, T \in \mathcal{L}(V)$. Prove that $ST = I$ if and only if $TS = I$.

Proof. Suppose first that $ST = I$. It is easy to see that defining $(ST)^{-1} = I^{-1} = I$ yields an inverse of ST. Thus, since ST is invertible, we have by Exercise 3.D.9 that S and T are invertible. It follows since $ST = I$ that $S = T^{-1}$, meaning that

$$TS = TT^{-1} = I$$

as desired.

The proof is symmetric in the other direction. ∎

3.E Products and Quotients of Vector Spaces

9/6:

- **Product** (of $V_1, \dots, V_m$): The set
$$V_1 \times \cdots \times V_m = \{(v_1, \dots, v_m) : v_1 \in V_1, \dots, v_m \in V_m\}$$
  - Addition on $V_1 \times \cdots \times V_m$ is defined by
$$(u_1, \dots, u_m) + (v_1, \dots, v_m) = (u_1 + v_1, \dots, u_m + v_m)$$
  - Scalar multiplication on $V_1 \times \cdots \times V_m$ is defined by
$$\lambda(v_1, \dots, v_m) = (\lambda v_1, \dots, \lambda v_m)$$
- The product of n vector spaces over $\mathbb{F}$ is a vector space over $\mathbb{F}$, with addition and scalar multiplication defined as above.
- We can, for example, identify $\mathbb{R}^2 \times \mathbb{R}^3$ with $\mathbb{R}^5$ by constructing an isomorphism from every vector $((x_1, x_2), (x_3, x_4, x_5)) \in \mathbb{R}^2 \times \mathbb{R}^3$ to the vector $(x_1, x_2, x_3, x_4, x_5) \in \mathbb{R}^5$.
- The dimension of the product of vector spaces is equal to the sum of the dimensions of the component vector spaces.

Theorem 3.20. *Suppose $V_1, \dots, V_m$ are finite-dimensional vector spaces. Then $V_1 \times \cdots \times V_m$ is finite-dimensional and*
$$\dim(V_1 \times \cdots \times V_m) = \dim V_1 + \cdots + \dim V_m$$

Proof. Choose a basis of each V_j. For each basis vector of each V_j, consider the element of $V_1 \times \cdots \times V_m$ that equals the basis vector in the j^{th} slot and 0 in the other slots. The list of all such vectors is linearly independent and spans $V_1 \times \cdots \times V_m$. Thus, it is a basis of $V_1 \times \cdots \times V_m$. The length of this basis is $\dim V_1 + \cdots + \dim V_m$, as desired. ■

- We now relate products and direct sums.

Theorem 3.21. *Suppose that $U_1, \dots, U_m$ are subspaces of V. Define a linear map $\Gamma : U_1 \times \cdots \times U_m \to U_1 + \cdots + U_m$ by*
$$\Gamma(u_1, \dots, u_m) = u_1 + \cdots + u_m$$
Then $U_1 + \cdots + U_m$ is a direct sum if and only if Γ is injective.

Proof. Suppose first that Γ is injective. Then the only way to write 0 as a sum $u_1 + \cdots + u_m$, where each $u_j \in U_j$, is by taking each $u_j = 0$. It follows by the condition on direct sums that $U_1 + \cdots + U_m$ is a direct sum. The proof is symmetric in the reverse direction. ■

  - Note that since Γ is surjective by the definition of $U_1 + \cdots + U_m$, the condition that Γ is injective could be changed to the condition that Γ is invertible.

- We can now prove that the dimensions add up in a direct sum.

Theorem 3.22. *Suppose V is finite-dimensional and $U_1, \dots, U_m$ are subspaces of V. Then $U_1 + \cdots + U_m$ is a direct sum if and only if*
$$\dim(U_1 + \cdots + U_m) = \dim U_1 + \cdots + \dim U_m$$

Proof. Suppose first that $U_1 + \cdots + U_m$ is a direct sum. Then by Theorem 3.21, there exists an invertible linear map Γ from $U_1 \times \cdots \times U_m$ to $U_1 + \cdots + U_m$. Thus, by Theorem 3.14, $U_1 \times \cdots \times U_m$ and $U_1 + \cdots + U_m$ have the same dimension. Therefore,
$$\begin{aligned}\dim(U_1 + \cdots + U_m) &= \dim(U_1 \times \cdots \times U_m) \\ &= \dim U_1 + \cdots + \dim U_m && \text{Theorem 3.20}\end{aligned}$$
as desired.

The proof is symmetric in the other direction. ■

- **Sum** (of $v \in V$ and U a subspace of V): The subset of V defined by

$$v + U = \{v + u : u \in U\}$$

- **Affine subset** (of V): A subset of V of the form $v + U$ for some $v \in V$ and some subspace U of V.
- **Parallel** (subset to U): An affine subset $v + U$ of V.
- **Quotient space**: The set of all affine subsets of V parallel to U.

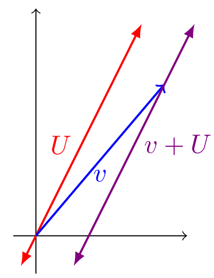

Figure 3.1: Visualizing $v + U$.

  - Symbolically,

$$V/U = \{v + U : v \in V\}$$

- Two affine subsets parallel to U are equal or disjoint.

Theorem 3.23. *Suppose U is a subspace of V and $v, w \in V$. Then the following are equivalent.*

(a) $v - w \in U$;

(b) $v + U = w + U$;

(c) $(v + U) \cap (w + U) \neq \emptyset$.

Proof. First, suppose that $v - w \in U$. Let $x \in v + U$ be arbitrary. Then $x = v + u$ for some $u \in U$. Now since $v - w \in U$, $u \in U$, and U is a subspace, we have that $v - w + u \in U$. Thus, $x = w - w + v + u = w + (v - w + u) \in w + U$. The proof is symmetric in the other direction. Therefore, $v + U = w + U$, as desired.

Second, suppose that $v + U = w + U$. Since U is nonempty ($0 \in U$ by definition), we know that $v + U \neq \emptyset \neq w + U$. Therefore, $(v + U) \cap (w + U) \supset \{0\} \neq \emptyset$, as desired.

Third, suppose that $(v + U) \cap (w + U) \neq \emptyset$. Then there exists x such that $x \in v + U$ and $x \in w + U$. It follows that $x = v + u_1$ and $x = w + u_2$ for some $u_1, u_2 \in U$. Thus, by transitivity, $v + u_1 = w + u_2$. Therefore, $v - w = u_2 - u_1 \in U$, as desired. ■

- **Sum** (of $v + U, w + U \in V/U$): The affine subset $(v + w) + U$.
- **Product** (of $v + U \in V/U$ and $\lambda \in \mathbb{F}$): The affine subset $(\lambda v) + U$.
- We now verify that the above operations are well-defined and prove that the quotient space is a vector space.

Theorem 3.24. *Suppose U is a subspace of V. Then V/U, with the operations of addition and scalar multiplication as defined above, is a vector space.*

Proof. The way affine subsets are defined, we may have $v + U = \hat{v} + U$ and yet have $v \neq \hat{v}$. Thus, we must first guarantee that the operations of addition and scalar multiplication, as defined above, are well-defined, i.e., that if $v + U = \hat{v} + U$ and $w + U = \hat{w} + U$, then $(v + w) + U = (\hat{v} + \hat{w}) + U$ and $(\lambda v) + U = (\lambda \hat{v}) + U$. Let's begin.

To confirm that addition as defined above is a well-defined operation, let $v, \hat{v}, w, \hat{w} \in V$ be such that $v + U = \hat{v} + U$ and $w + U = \hat{w} + U$. Then by Theorem 3.23, $v - \hat{v} \in U$ and $w - \hat{w} \in U$. It follows since U is a subspace that $(v - \hat{v}) + (w - \hat{w}) \in U$. Consequently, $(v + w) - (\hat{v} + \hat{w}) \in U$, so by Theorem 3.23 again, $(v + w) + U = (\hat{v} + \hat{w}) + U$, as desired.

Similarly, $v + U = \hat{v} + U$ implies $v - \hat{v} \in U$, implies $\lambda(v - \hat{v}) = \lambda v - \lambda \hat{v} \in U$, implies $(\lambda v) + U = (\lambda \hat{v}) + U$, as desired.

The remaining proof that V/U is a vector space is straightforward; note that $0 + U$ is the identity element and $(-v) + U$ is the additive inverse. ■

- **Quotient map**: The linear map $\pi : V \to V/U$ defined by $\pi(v) = v + U$ for all $v \in V$.
- We now give a formula for the dimension of a quotient space.

Theorem 3.25. *Suppose V is finite-dimensional and U is a subspace of V. Then*

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

Proof. Let π be the quotient map from V to V/U. From Theorem 3.23, we know that in order for $w + U = 0 + U$, we must have $v - 0 = v \in U$. Thus, $\pi(u) = 0$ if and only if $u \in U$, meaning $\operatorname{null} \pi = U$. Additionally, we clearly have that $\operatorname{range} \pi = V/U$. Therefore, by the Fundamental Theorem of Linear Maps, we have that

$$\begin{aligned}
\dim V &= \dim \operatorname{null} \pi + \dim \operatorname{range} \pi \\
&= \dim U + \dim V/U \\
\dim V/U &= \dim V - \dim U
\end{aligned}$$

as desired. ■

- Lastly, consider the fact that we can add any vector in the null space of a linear map T to an argument passed to T without changing its output. In other words, if $T \in \mathcal{L}(V, W)$, $v \in V$, and $u \in \operatorname{null} T$, then $T(v + u) = Tv + Tu = Tv$. We formalize this concept with the following definition.
- $\boldsymbol{\tilde{T}}$: The function from $V/(\operatorname{null} T)$ to W defined by $\tilde{T}(v + \operatorname{null} T) = Tv$, where $T \in \mathcal{L}(V, W)$.
- We now state a few basic results about $\tilde{T}$.

Theorem 3.26. *Suppose $T \in \mathcal{L}(V, W)$. Then*

(*a*) *$\tilde{T}$ is a linear map from $V/(\operatorname{null} T)$ to W;*

(*b*) *$\tilde{T}$ is injective;*

(*c*) $\operatorname{range} \tilde{T} = \operatorname{range} T$*;*

(*d*) *$V/(\operatorname{null} T)$ is isomorphic to* $\operatorname{range} T$.

3.F Duality

9/7: - **Linear functional** (on v): A linear map from V to $\mathbb{F}$.
  - In other words, a linear functional is an element of $\mathcal{L}(V, \mathbb{F})$.

- **Dual space** (of V): The vector space of all linear functionals on V. *Denoted by* $\boldsymbol{V'}$. *Also known as* $\boldsymbol{V^*}$. *Given by*
$$V' = \mathcal{L}(V, \mathbb{F})$$
- We now give a definition of the dimension of the dual space.

Theorem 3.27. *Suppose V is finite-dimensional. Then V' is also finite-dimensional and*
$$\dim V' = \dim V$$

Proof. By Theorem 3.16, we have that
$$\begin{aligned}\dim V' &= \dim \mathcal{L}(V, \mathbb{F})\\ &= (\dim V)(\dim \mathbb{F})\\ &= (\dim V)(1)\\ &= \dim V\end{aligned}$$
as desired. ■

- **Dual basis** (of a basis $v_1, \dots, v_n$ of V): The list $\varphi_1, \dots, \varphi_n$ of elements of V', where each φ_j is the linear functional on V such that
$$\varphi_j(v_k) = \begin{cases} 1 & k = j\\ 0 & k \neq j\end{cases}$$
where $v_1, \dots, v_n$ is a basis of V.
- We now verify that the dual basis of a basis of V is actually a basis of the dual space.

Theorem 3.28. *Suppose V is finite-dimensional. Then the dual basis of a basis of V is a basis of V'.*

Proof. Let $v_1, \dots, v_n$ be a basis of V, and let $\varphi_1, \dots, \varphi_n$ be the corresponding dual basis. Since the dual basis has length equal to the dimension of V' (by Theorem 3.27), Theorem 2.12 tells us that it will suffice to show that $\varphi_1, \dots, \varphi_n$ is linearly independent to confirm that it is a basis of V'. To do so, suppose
$$a_1\varphi_1 + \cdots + a_n\varphi_n = 0$$
where $a_1, \dots, a_n \in \mathbb{F}$ and 0 denotes the zero transformation. Since $(a_1\varphi_1 + \cdots + a_n\varphi_n)(v_j) = a_j$ for $j = 1, \dots, n$, we have that for any vector $c_1v_1 + \cdots + c_nv_n \in V$,
$$(a_1\varphi_1 + \cdots + a_n\varphi_n)(c_1v_1 + \cdots + c_nv_n) = c_1a_1 + \cdots + c_na_n$$
Therefore, the only way to guarantee that $c_1a_1 + \cdots + c_na_n = 0$ is to let $a_1 = \cdots = a_n = 0$, as desired. ■

- **Dual map** (of $T \in \mathcal{L}(V, W)$): The linear map $T' \in \mathcal{L}(W', V')$ defined by
$$T'(\varphi) = \varphi \circ T$$
for all $\varphi \in W'$. *Also known as* $\boldsymbol{T^*}$.
- We now prove some algebraic properties of dual maps.

Theorem 3.29.

(a) $(S + T)' = S' + T'$ *for all* $S, T \in \mathcal{L}(V, W)$.

Proof. Let $S,T \in \mathcal{L}(V,W)$ be arbitrary. To prove that $(S+T)' = S'+T'$, it will suffice to show that $(S+T)'(\varphi) = (S'+T')(\varphi)$ for all $\varphi \in W'$. Let $\varphi \in W'$ be arbitrary. However, before we go into the main equality, it will be useful if we verify that $\varphi \circ (S+T) = \varphi \circ S + \varphi \circ T$. To do so, it will suffice to show that $(\varphi \circ (S+T))(v) = (\varphi \circ S + \varphi \circ T)(v)$ for all $v \in V$. Let $v \in V$ be arbitrary. Then

$$\begin{aligned}
(\varphi \circ (S+T))(v) &= \varphi((S+T)(v)) \\
&= \varphi(S(v)+T(v)) \\
&= \varphi(S(v)) + \varphi(T(v)) \\
&= (\varphi \circ S)(v) + (\varphi \circ T)(v) \\
&= (\varphi \circ S + \varphi \circ T)(v)
\end{aligned}$$

Now we can show that

$$\begin{aligned}
(S+T)'(\varphi) &= \varphi \circ (S+T) \\
&= \varphi \circ S + \varphi \circ T \\
&= S'(\varphi) + T'(\varphi) \\
&= (S'+T')(\varphi)
\end{aligned}$$

as desired. ■

(*b*) $(\lambda T)' = \lambda T'$ *for all* $\lambda \in \mathbb{F}$ *and all* $T \in \mathcal{L}(V,W)$.

Proof. The proof is symmetric to the proof of part (a). ■

(*c*) $(ST)' = T'S'$ *for all* $T \in \mathcal{L}(U,V)$ *and* $S \in \mathcal{L}(V,W)$.

Proof. Let $\varphi \in W'$ be arbitrary. Then

$$(ST)'(\varphi) = \varphi \circ (ST) = (\varphi \circ S) \circ T = T'(\varphi \circ S) = T'(S'(\varphi)) = (T'S')(\varphi)$$

as desired. ■

- **Annihilator** (of $U \subset V$): The set

$$U^0 = \{\varphi \in V' : \varphi(u) = 0 \ \forall \ u \in U\}$$

- The annihilator is a subspace.

Theorem 3.30. *Suppose $U \subset V$. Then U^0 is a subspace of V'*

Proof. To prove that U^0 is a subspace of V', it will suffice to show that $0 \in U^0$, $\varphi, \psi \in U^0$ implies $\varphi + \psi \in U^0$, and $\varphi \in U^0$ and $\lambda \in \mathbb{F}$ imply $\lambda\varphi \in U^0$. Let's begin.

Since $0(u) = 0$ for all $u \in U$, $0 \in U^0$.

Let $\varphi, \psi \in U^0$ be arbitrary. Let $u \in U$ be arbitrary. Then $(\varphi+\psi)(u) = \varphi(u) + \psi(u) = 0 + 0 = 0$, as desired.

The proof is symmetric for scalar multiplication. ■

- Dimension of the annihilator.

Theorem 3.31. *Suppose V is finite-dimensional and U is a subspace of V. Then*

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

Proof. Let $i \in \mathcal{L}(U,V)$ be the identity map $i(u) = u$ for all $u \in U$. Then $i' : V' \to U'$ is a linear map. It follows by the Fundamental Theorem of Linear Maps that

$$\dim \operatorname{range} i' + \dim \operatorname{null} i' = \dim V'$$

Since $i'(\varphi) = \varphi \circ i = \varphi$ for all $\varphi \in V'$, and $U^0 = \{\varphi \in V' : \varphi = 0\}$, we have that $i'(\varphi) = 0$ for all $\varphi \in U^0$. Thus, $U^0 = \operatorname{null} i'$. Additionally, we have that $\dim V = \dim V'$ by Theorem 3.27. Lastly, let $\psi \in U'$ be arbitrary. Define $\psi \in V'$ by

$$\psi(v) = \begin{cases} \varphi(v) & v \in U \\ 0 & v \notin U \end{cases}$$

Thus, $i'(\psi) = \psi \circ i = \varphi$. It follows that $\varphi \in \operatorname{range} i'$. Consequently, $\operatorname{range} i' = U'$, so $\dim U = \dim U' = \dim \operatorname{range} i'$ by Theorem 3.27. Therefore, we have from the first equation and the three substitutions that

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

as desired.[1] ■

- We now describe the null space of T'.

  Theorem 3.32. *Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V,W)$. Then*

  (*a*) $\operatorname{null} T' = (\operatorname{range} T)^0$.

  Proof. First, let $\varphi \in \operatorname{null} T'$ be arbitrary. Then $T'(\varphi) = \varphi \circ T = 0$. It follows that $0 = (\varphi \circ T)(v) = \varphi(Tv)$ for all $v \in V$. But this means that φ is a linear functional that maps every element of $\operatorname{range} T$ to 0, i.e., that $\varphi \in (\operatorname{range} T)^0$. The proof is symmetric in the other direction. ■

  (*b*) $\dim \operatorname{null} T' = \dim \operatorname{null} T + \dim W - \dim V$.

  Proof. We have that

$$\begin{aligned} \dim \operatorname{null} T' &= \dim (\operatorname{range} T)^0 && \text{Theorem 3.32a} \\ &= \dim W - \dim \operatorname{range} T && \text{Theorem 3.31} \\ &= \dim W - (\dim V - \dim \operatorname{null} T) && \text{Fundamental Theorem of Linear Maps} \\ &= \dim \operatorname{null} T + \dim W - \dim V \end{aligned}$$

  as desired. ■

  - Note that the proof of part (a) does not use the hypothesis that V, W are finite-dimensional, so the argument holds for infinite-dimensional vector spaces as well.

- T surjective is equivalent to T' injective.

  Theorem 3.33. *Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V,W)$. Then T is surjective if and only if T' is injective.*

  Proof. Suppose first that T is surjective. Then $\operatorname{range} T = W$. It follows by Theorem 3.31 that

$$\dim(\operatorname{range} T)^0 = \dim W - \dim \operatorname{range} T = 0$$

  meaning that $(\operatorname{range} T)^0 = \{0\}$. Thus, by Theorem 3.32a, $\operatorname{null} T' = \{0\}$. Therefore, by Theorem 3.4, T' is injective, as desired.

  The proof is symmetric in the other direction. ■

[1] Note that we may also prove this by constructing a basis of U extending it to a basis of V, and showing that the extended portion of the dual basis is a basis of U^0.

- We now describe the range space of T'.

Theorem 3.34. *Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V,W)$. Then*

(*a*) $\dim \operatorname{range} T' = \dim \operatorname{range} T$.

Proof. We have that

$$\begin{aligned}
\dim \operatorname{range} T' &= \dim W' - \dim \operatorname{null} T' && \text{Fundamental Theorem of Linear Maps} \\
&= \dim W - \dim \operatorname{null} T' && \text{Theorem 3.27} \\
&= \dim W - \dim (\operatorname{range} T)^0 && \text{Theorem 3.32a} \\
&= \dim \operatorname{range} T && \text{Theorem 3.31}
\end{aligned}$$

as desired. ■

(*b*) $\operatorname{range} T' = (\operatorname{null} T)^0$.

Proof. First, let $\varphi \in \operatorname{range} T'$ be arbitrary. Then there exists $\psi \in W'$ such that $\varphi = T'(\psi)$. Now let $v \in \operatorname{null} T$ be arbitrary. It follows that

$$\varphi(v) = (T'(\psi))(v) = (\psi \circ T)(v) = \psi(Tv) = \psi(0) = 0$$

Therefore, $\varphi \in (\operatorname{null} T)^0$, as desired.
Second, we have that

$$\begin{aligned}
\dim \operatorname{range} T' &= \dim \operatorname{range} T && \text{Theorem 3.34a} \\
&= \dim V - \dim \operatorname{null} T && \text{Fundamental Theorem of Linear Maps} \\
&= \dim (\operatorname{null} T)^0 && \text{Theorem 3.31}
\end{aligned}$$

Therefore, since Theorem 3.5 implies that $\operatorname{range} T'$ is a subspace of $(\operatorname{null} T)^0$ and $\dim \operatorname{range} T' = \dim (\operatorname{null} T)^0$, Exercise 2.C.1 asserts that $\operatorname{range} T' = (\operatorname{null} T)^0$, as desired. ■

- T injective is equivalent to T' surjective.

Theorem 3.35. *Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V,W)$. Then T is injective if and only if T' is surjective.*

Proof. Suppose first that T is injective. Then by Theorem 3.4, $\operatorname{null} T = \{0\}$. Thus, since Theorem 3.2 asserts that $\varphi(0) = 0$ for any linear functional, we have that every linear functional is in the annihilator of $\operatorname{null} T$, i.e., that $(\operatorname{null} T)^0 = V'$. It follows by Theorem 3.34b that $\operatorname{range} T' = V'$. Therefore, T' is surjective, as desired.

The proof is symmetric in the other direction. ■

9/8:

- **Transpose** (of an m-by-n matrix A): The matrix obtained from A by interchanging the rows and columns. More specifically, the n-by-m matrix A^t whose entries are given by $(A^t)_{k,j} = A_{j,k}$. *Denoted by* $\boldsymbol{A^t}$.
- Properties of the transpose:

$$(A + C)^t = A^t + C^t \qquad\qquad (\lambda A)^t = \lambda A^t$$

- Transpose of a product.

Theorem 3.36. *If A is an m-by-n matrix and C is an n-by-p matrix, then*

$$(AC)^t = C^t A^t$$

Proof. We have that

$$\begin{aligned}((AC)^t)_{k,j} &= (AC)_{j,k} \\ &= \sum_{r=1}^{n} A_{j,r}C_{r,k} \\ &= \sum_{r=1}^{n} (C^t)_{k,r}(A^t)_{r,j} \\ &= (C^tA^t)_{k,j}\end{aligned}$$

for all $1 \leq k \leq p$ and $1 \leq j \leq m$, as desired. ■

- We now show that the transpose and the dual map are essentially the same object.

Theorem 3.37. *Suppose $T \in \mathcal{L}(V,W)$. Then $\mathcal{M}(T') = (\mathcal{M}(T))^t$.*

Proof. Let $v_1, \dots, v_n$ be a basis of V, and let $\varphi_1, \dots, \varphi_n$ be the corresponding dual basis of V'. Similarly, let $w_1, \dots, w_m$ be a basis of W, and let $\psi_1, \dots, \psi_m$ be the corresponding dual basis of W'. Let $A = \mathcal{M}(T)$ and $C = \mathcal{M}(T')$. Let $1 \leq j \leq m$ and $1 \leq k \leq n$ be arbitrary. Then we have from the definition of $\mathcal{M}(T')$ that

$$T'(\psi_j) = \sum_{r=1}^{n} C_{r,j}\varphi_r$$

from the definition of T' that

$$\begin{aligned}(\psi \circ T)(v_k) &= \sum_{r=1}^{n} C_{r,j}\varphi_r(v_k) \\ &= C_{k,j}\end{aligned}$$

and from the definition of $\mathcal{M}(T)$ that

$$\begin{aligned}(\psi \circ T)(v_k) &= \psi_j(Tv_k) \\ &= \psi_j\left(\sum_{r=1}^{m} A_{r,k}w_r\right) \\ &= \sum_{r=1}^{m} A_{r,k}\psi_j(w_r) \\ &= A_{j,k}\end{aligned}$$

Therefore, from the last two results, we have by transitivity that $A_{j,k} = C_{k,j}$ for all $1 \leq j \leq m$ and $1 \leq k \leq n$. It follows that $C = A^t$, i.e., that $\mathcal{M}(T') = (\mathcal{M}(T))^t$, as desired. ■

- **Row rank** (of a matrix A): The dimension of the span of the rows of A in $\mathbb{F}^{1,n}$.
- **Column rank** (of a matrix A): The dimension of the span of the columns of A in $\mathbb{F}^{m,1}$.
- The dimension of range T equals the column rank of $\mathcal{M}(T)$.

Theorem 3.38. *Suppose V and W are finite-dimensional and $T \in \mathcal{L}(V,W)$. Then* $\dim \operatorname{range} T$ *equals the column rank of $\mathcal{M}(T)$.*

Proof. Let $v_1, \dots, v_n$ be a basis of V, and let $w_1, \dots, w_m$ be a basis of W. Since $Tv = c_1Tv_1 + \cdots + c_nTv_n$ for all $Tv \in \operatorname{range} T$ (because $v = c_1v_1 + \cdots + c_nTv_n$ for some $c_1, \dots, c_n \in \mathbb{F}$ for all $v \in V$, and T is a linear map), we have that $\operatorname{range} T = \operatorname{span}(Tv_1, \dots, Tv_n)$. Additionally, since $\mathcal{M}$ is an isomorphism from $\operatorname{span}(Tv_1, \dots, Tv_n)$ to $\operatorname{span}(\mathcal{M}(Tv_1), \dots, \mathcal{M}(Tv_n))$, Theorem 3.14 asserts that $\dim\operatorname{span}(Tv_1, \dots, Tv_n) = \dim\operatorname{span}(\mathcal{M}(Tv_1), \dots, \mathcal{M}(Tv_n))$. Therefore,

$$\begin{aligned}\dim\operatorname{range} T &= \dim\operatorname{span}(Tv_1, \dots, Tv_n) \\ &= \dim\operatorname{span}(\mathcal{M}(Tv_1), \dots, \mathcal{M}(Tv_n))\end{aligned}$$

where the latter value is the column rank, as desired. ■

- Row rank equals column rank.

Theorem 3.39. *Suppose $A \in \mathbb{F}^{m,n}$. Then the row rank of A equals the column rank of A.*

Proof. Let $T : \mathbb{F}^{n,1} \to \mathbb{F}^{m,1}$ be defined by $Tx = Ax$. It follows that $\mathcal{M}(T) = A$. Thus,

$$\begin{aligned}\text{column rank } A &= \text{column rank } \mathcal{M}(T) \\ &= \dim\operatorname{range} T && \text{Theorem 3.38} \\ &= \dim\operatorname{range} T' && \text{Theorem 3.34a} \\ &= \text{column rank } \mathcal{M}(T') && \text{Theorem 3.38} \\ &= \text{column rank } A^t && \text{Theorem 3.37} \\ &= \text{row rank } A\end{aligned}$$

as desired. ■

- **Rank** (of A): The column rank of A.

Chapter 4

Polynomials

9/8:

- **Real part** (of $a + bi \in \mathbb{C}$): The number a. *Denoted by* $\mathbf{Re}\,\boldsymbol{z}$.
- **Imaginary part** (of $a + bi \in \mathbb{C}$): The number b. *Denoted by* $\mathbf{Re}\,\boldsymbol{z}$.
- **Complex conjugate** (of $z \in \mathbb{C}$): The number $\operatorname{Re} z - (\operatorname{Im} z)i$. *Denoted by* $\bar{\boldsymbol{z}}$.
- **Absolute value** (of $z \in \mathbb{C}$): The number $\sqrt{(\operatorname{Re} z)^2 + (\operatorname{Im} z)^2}$. *Denoted by* $|\boldsymbol{z}|$.
- $z = \bar{z}$ if and only if $z \in \mathbb{R}$.
- Properties of complex numbers.

  Theorem 4.1. *Suppose $w, z \in \mathbb{C}$. Then*

  sum of z and $\bar{z}$
  $z + \bar{z} = 2 \operatorname{Re} z$.

  difference of z and $\bar{z}$
  $z - \bar{z} = 2(\operatorname{Im} z)i$.

  product of z and $\bar{z}$
  $z\bar{z} = |z|^2$.

  additivity and multiplicativity of the complex conjugate
  $\overline{w + z} = \bar{w} + \bar{z}$ *and* $\overline{wz} = \bar{w}\bar{z}$.

  conjugate of conjugate
  $\bar{\bar{z}} = z$.

  real and imaginary parts are bounded by $|z|$
  $|\operatorname{Re} z| \leq |z|$ *and* $|\operatorname{Im} z| \leq |z|$.

  absolute value of the complex conjugate
  $|\bar{z}| = |z|$.

  multiplicativity of absolute value
  $|wz| = |w|\,|z|$.

  Triangle Inequality
  $|w + z| \leq |w| + |z|$.

- If a polynomial is the zero function, then all coefficients are 0.
  - It follows that the coefficients of a polynomial are uniquely determined.
- **Division Algorithm** (for integers): If p, s are nonnegative integers with $s \neq 0$, then there exist nonnegative integers q, r such that $r < s$ and

$$p = sq + r$$

- Analogously,

  Theorem 4.2 (Division Algorithm for Polynomials)**.** *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 $m=\deg s$. We divide into two cases ($n<m$ and $n\geq m$). If $n<m$, then take $q=0$ and $r=p$.

  On the other hand, if $n\geq P$, then let $T:\mathcal{P}_{n-m}(\mathbb{F})\times\mathcal{P}_{m-1}(\mathbb{F})\to\mathcal{P}_n(()\mathbb{F})$ be defined by

  $$T(q,r)=sq+r$$

  We can easily confirm that T is a linear map.

  We now seek to prove that $\operatorname{null} T=\{(0,0)\}$. Let $(q,r)\in\operatorname{null} T$ be arbitrary. Then $sq+r=0$. It follows that all coefficients of the polynomial $sq+r$ are zero. Consequently, $q=0$ and $r=0$, as desired. Therefore, $\dim\operatorname{null} T=0$. Additionally, Theorem 3.20 implies that

  $$\dim(\mathcal{P}_{n-m}(\mathbb{F})\times\mathcal{P}_{m-1}(\mathbb{F}))=(n-m+1)+(m-1+1)=n+1$$

  It follows by the Fundamental Theorem of Linear Maps that

  $$\begin{aligned}\dim\operatorname{range} T&=\dim(\mathcal{P}_{n-m}(\mathbb{F})\times\mathcal{P}_{m-1}(\mathbb{F}))-\dim\operatorname{null} T\\&=n+1\\&=\dim\mathcal{P}_n(\mathbb{F})\end{aligned}$$

  Thus, by Exercise 2.C.1, $\operatorname{range} T=\mathcal{P}_n(\mathbb{F})$. Therefore, since $p\in\mathcal{P}_n(\mathbb{F})$, we know that there exists $q\in\mathcal{P}_{n-m}(\mathbb{F})$ and $r\in\mathcal{P}_{m-1}(\mathbb{F})$ such that $p=T(q,r)=sq+r$.

  Additionally, we know that q,r are unique: If there exist q',r' such that $T(q',r')=p$, then $T(q-q',r-r')=p-p=0$, implying since $\operatorname{null} T=\{(0,0)\}$ that $q-q'=0$ and $r-r'=0$, i.e., that $q=q'$ and $r=r'$. ■

- **Zero** (of $p\in\mathcal{P}(\mathbb{F})$): A number $\lambda\in\mathbb{F}$ such that $p(\lambda)=0$. *Also known as* **root**.
- **Factor** (of $p\in\mathcal{P}(\mathbb{F})$): A polynomial $s\in\mathcal{P}(\mathbb{F})$ such that there exists a polynomial $q\in\mathcal{P}(\mathbb{F})$ satisfying $p=sq$.
- We now relate zeroes and factors.

  Theorem 4.3. *Suppose $p\in\mathcal{P}(\mathbb{F})$ and $\lambda\in\mathbb{F}$. Then $p(\lambda)=0$ if and only if there is a polynomial $q\in\mathcal{P}(\mathbb{F})$ such that*

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

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

- Putting bounds on the number of zeroes a polynomial can have.

  Theorem 4.4. *Suppose $p\in\mathcal{P}(\mathbb{F})$ is a polynomial with degree $m\geq 0$. Then p has at most m distinct zeros in $\mathbb{F}$.*

- We cannot prove the following without complex analysis, but we will state it, regardless.

  Theorem 4.5 (Fundamental Theorem of Algebra)**.** *Every nonconstant polynomial with complex coefficients has a zero.*

- The following proceeds immediately from the Fundamental Theorem of Algebra.

  Theorem 4.6. *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 $c, \lambda_1, \dots, \lambda_m \in \mathbb{C}$.

- We now explore some of the differences between $\mathbb{R}$ and $\mathbb{C}$.

  Theorem 4.7. *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 $\bar{\lambda}$.*

  Theorem 4.8. *Suppose $b, c \in \mathbb{R}$. Then there is a polynomial factorization of the form*

  $$x^2 + bx + c = (x - \lambda_1)(x - \lambda_2)$$

  with $\lambda_1, \lambda_2 \in \mathbb{R}$ if and only if $b^2 \geq 4c$.

  Theorem 4.9. *Suppose $p \in \mathcal{P}(\mathbb{R})$ is a nonconstant polynomial. Then p has a unique factorization (except for the order of 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 $c, \lambda_1, \dots, \lambda_m, b_1, \dots, b_M, c_1, \dots, c_M \in \mathbb{R}$, with $b_j^2 < 4c_j$ for each j.

Chapter 5

Eigenvalues, Eigenvectors, and Invariant Subspaces

5.A Invariant Subspaces

9/8:

- Let $T \in \mathcal{L}(V)$, and let V be decomposable into a direct sum of proper subspaces as follows.

$$V = U_1 \oplus \cdots \oplus U_m$$

  - To understand T, we need only understand each each **restriction** of T to a U_j.
  - Since $T|_{U_j}$ may not map U_j onto itself in every case, to use operator-based tools, we need to consider only direct sum decompositions into subspaces that T maps onto themselves, or **invariant subspace**.
- **Invariant subspace** (of V under T): A subspace U of V such that $u \in U$ implies $Tu \in U$, where $T \in \mathcal{L}(V)$.
  - In other words, U is invariant under T iff $T|_U \in \mathcal{L}(U)$.
- Some invariant subspaces under $T \in \mathcal{L}(V)$: $\{0\}$, V, null T, and range T.
- **Invariant subspace problem**: The most famous unsolved problem in functional analysis, dealing with invariant subspaces of operators on infinite-dimensional vector spaces.
- To begin our study of invariant subspaces, we consider the simplest possible type of invariant subspace: those with dimension 1.
- Every 1-dimensional subspace of V is of the form span(v) for some $v \in V$.
  - If span(v) is invariant under $T \in \mathcal{L}(V)$, then $Tv \in$ span(v).
  - If $Tv \in$ span(v), then there exists $\lambda \in \mathbb{F}$ such that $Tv = \lambda v$.
- **Eigenvalue** (of T): A number $\lambda \in \mathbb{F}$ such that there exists a nonzero vector $v \in V$ satisfying the equation $Tv = \lambda v$. *Also known as* **characteristic value**.
- "T has a 1-dimensional invariant subspace if and only if T has an eigenvalue" (Axler, 2015, p. 134).
- We now give some conditions λ can satisfy to be deemed an eigenvalue.

  Theorem 5.1. *Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, $I \in \mathcal{L}(V)$ is the identity operator on V, and $\lambda \in \mathbb{F}$. Then the following are equivalent.*

  (a) λ 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.*

Proof. Suppose first that λ is an eigenvalue of T. Then

$$\begin{aligned} Tv &= \lambda v \\ Tv &= \lambda I v \\ Tv - \lambda I v &= 0 \\ (T - \lambda I)v &= 0 \end{aligned}$$

for some $v \in V$ such that $v \neq 0$. It follows that $v \in \operatorname{null}(T - \lambda I)$, so by Theorem 3.4, $T - \lambda I$ is not injective, as desired. The proof is symmetric in the other direction. Therefore, conditions (a) and (b) are equivalent.

To prove that (a), (b), (c), and (d) are equivalent at this point, it will suffice to show that (b), (c), and (d) are equivalent. But we have this by Theorem 3.13, as desired. ■

- **Eigenvector** (of T): A nonzero vector $v \in V$ such that there exists a $\lambda \in \mathbb{F}$ satisfying the equation $Tv = \lambda v$.
- Since $Tv = \lambda v$ iff $(T - \lambda I)v = 0$, "a vector $v \in V$ with $v \neq 0$ is an eigenvector of T corresponding to λ if and only if $v \in \operatorname{null}(T - \lambda I)$" (Axler, 2015, p. 135).
- Eigenvectors corresponding to distinct eigenvalues are linearly independent.

Theorem 5.2. *Let $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \dots, \lambda_m$ are distinct eigenvalues of T and $v_1, \dots, v_m$ are corresponding eigenvectors. Then $v_1, \dots, v_m$ is linearly independent.*

Proof. Suppose for the sake of contradiction that $v_1, \dots, v_m$ is linearly dependent. Then by the Linear Dependence Lemma, we may let k be the smallest positive integer such that $v_k \in \operatorname{span}(v_1, \dots, v_{k-1})$. It follows that

$$v_k = a_1 v_1 + \cdots + a_{k-1} v_{k-1}$$

for some $a_1, \dots, a_{k-1} \in \mathbb{F}$. Thus, applying T, we have that

$$\begin{aligned} Tv_k &= a_1 T v_1 + \cdots + a_{k-1} T v_{k-1} \\ \lambda_k v_k &= a_1 \lambda_1 v_1 + \cdots + a_{k-1} \lambda_{k-1} v_{k-1} \end{aligned}$$

If we multiply the first equation by λ_k and subtract the above equation from it, we get that

$$0 = a_1(\lambda_k - \lambda_1)v_1 + \cdots + a_{k-1}(\lambda_k - \lambda_{k-1})v_{k-1}$$

But since k is the smallest positive integer j such that $v_j \in \operatorname{span}(v_1, \dots, v_{j-1})$, we know that $v_1, \dots, v_{k-1}$ are linearly independent. Thus, $a_1(\lambda_k - \lambda_1) = \cdots = a_{k-1}(\lambda_k - \lambda_{k-1}) = 0$. More specifically, since all eigenvalues are distinct (i.e., $\lambda_k - \lambda_j \neq 0$ for any $j = 1, \dots, k-1$), we must have that $a_1 = \cdots = a_{k-1} = 0$. But this implies that

$$\begin{aligned} v_k &= a_1 v_1 + \cdots + a_{k-1} v_{k-1} \\ &= 0 \end{aligned}$$

contradicting the fact that v_k, as an eigenvector, is nonzero. ■

- We now put a bound on the number of eigenvalues.

Theorem 5.3. *Suppose V is finite-dimensional. Then each operator on V has at most* $\dim V$ *distinct eigenvalues.*

Proof. Let $T \in \mathcal{L}(V)$ have distinct eigenvalues $\lambda_1, \dots, \lambda_m$ and corresponding eigenvectors $v_1, \dots, v_m$. Then by Theorem 5.2, $v_1, \dots, v_m$ is linearly independent. It follows by Theorem 2.3 that $m \leq \dim V$ ■

- **Restriction operator** (of $T : V \to W$ to $U \subset V$): The function $T|_U : U \to W$ defined by $T|_U(u) = Tu$ for all $u \in U$. *Denoted by* $\boldsymbol{T|_U}$.
  - The fact that U is invariant under T is what allows us to consider $T|_U$ to be in $\mathcal{L}(U)$ as opposed to just $\mathcal{L}(V)$.
- **Quotient operator**: The operator $T/U \in \mathcal{L}(V/U)$ defined by $(T/U)(v + U) = Tv + U$ for all $v \in V$.
- Axler (2015) verifies that the restriction operator and the quotient operator actually *are* operators, in general, as defined.

5.B Eigenvectors and Upper-Triangular Matrices

9/10:

- If an operator $T \in \mathcal{L}(V)$, then $TT = T^2 \in \mathcal{L}(V)$.
- $\boldsymbol{T^m}$: The operator $T^m \in \mathcal{L}(V)$ defined by

$$T^m = \underbrace{T \cdots T}_{m \text{ times}}$$

  where $T \in \mathcal{L}(V)$, $m \in \mathbb{N}$.
- $\boldsymbol{T^0}$: The identity operator $I \in \mathcal{L}(V)$, where $T \in \mathcal{L}(V)$.
- $\boldsymbol{T^{-m}}$: The operator $T^{-m} \in \mathcal{L}(V)$ defined by

$$T^{-m} = (T^{-1})^m$$

  where $T \in \mathcal{L}(V)$ is invertible with inverse T^{-1}, and $m \in \mathbb{N}$.
- It follows from these definitions that

$$T^m T^n = T^{m+n} \qquad\qquad (T^m)^n = T^{mn}$$

  for any $m, n \in \mathbb{Z}$ if T is invertible and for any $m, n \in \mathbb{N}$ if T is not invertible.
- $\boldsymbol{p(T)}$: The operator defined by

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

  where $T \in \mathcal{L}(V)$, and $p \in \mathcal{P}(\mathbb{F})$ is defined by $p(z) = a_0 + a_1 z + \cdots + a_m z^m$ for all $z \in \mathbb{F}$.
- $f : \mathcal{P}(\mathbb{F}) \to \mathcal{L}(V)$ defined by $p \mapsto p(T)$ is linear.
- **Product** (of $p, q \in \mathcal{P}(\mathbb{F})$): The polynomial $pq \in \mathcal{P}(\mathbb{F})$ defined by $(pq)(z) = p(z)q(z)$ for all $z \in \mathbb{F}$.
- Multiplicative properties of $p(T)$.

Theorem 5.4. *Suppose $p, q \in \mathcal{P}(\mathbb{F})$ and $T \in \mathcal{L}(V)$. Then*

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

Proof. Suppose $p(z) = \sum_{j=0}^{m} a_j z^j$ and $q(z) = \sum_{k=0}^{n} b_k z^k$ for all $z \in \mathbb{F}$. Then

$$(pq)(z) = \sum_{j=0}^{m} \sum_{k=0}^{n} a_j b_k z^{j+k}$$

so

$$\begin{aligned}(pq)(T) &= \sum_{j=0}^{m}\sum_{k=0}^{n} a_j b_k T^{j+k} \\ &= \left(\sum_{j=0}^{m} a_j T^j\right)\left(\sum_{k=0}^{n} b_k T^k\right) \\ &= p(T)q(T)\end{aligned}$$

as desired. ∎

(*b*) $p(T)q(T) = q(T)p(T)$.

Proof. It follows from part (a) that $p(T)q(T) = (pq)(T) = (qp)(T) = q(T)p(T)$, as desired. ∎

- We now prove a central result concerning operators on complex vector spaces.

  Theorem 5.5. *Every operator on a finite-dimensional, nonzero, complex vector space has an eigenvalue.*

  Proof. Let V be nonzero complex vector space of dimension n, let $T \in \mathcal{L}(V)$, and let $v \in V$ be nonzero. Then since $\operatorname{len}(v, Tv, T^2v, \dots, T^nv) > \dim V$, Theorem 2.3 implies that $v, Tv, \dots, T^nv$ is not linearly independent. Thus, there exist $a_0, \dots, a_n \in \mathbb{F}$, not all 0, such that

  $$0 = a_0v + a_1Tv + \cdots + a_nT^nv$$

  Additionally, we have that $a_1, \dots, a_n$ do not all equal zero (suppose they did; then $0 = a_0v$, so $a_0 = 0$ because $v \neq 0$; this would imply that $a_0 = \cdots = a_n = 0$, a contradiction). Thus, if we consider the (nonconstant, by the previous result) polynomial with coefficients equal to the a's, Theorem 4.6 implies that

  $$a_0 + a_1z + \cdots + a_nz^n = c(z - \lambda_1)\cdots(z - \lambda_m)$$

  for all $z \in \mathbb{C}$, where $c \in \mathbb{C}$ is nonzero and each $\lambda_j \in \mathbb{C}$[1]. It follows that

  $$\begin{aligned}0 &= a_0v + a_1Tv + \cdots + a_nT^nv \\ &= (a_0I + a_1T + \cdots + a_nT^n)v \\ &= c(T - \lambda_1I)\cdots(T - \lambda_mI)v && \text{Theorem 5.4}\end{aligned}$$

  Each $T - \lambda_jI$ is a linear operator in its own right. As a linear map, we have by Theorem 3.2 that $(T - \lambda_jI)0 = 0$ for all $j = 1, \dots, m$. However, $v \neq 0$, meaning that there exists $1 \leq k \leq m$ such that $T - \lambda_kI$ maps its nonzero argument, which we may call v' and know to be the result of all of the operators to its right being applied successively to v via function composition, to 0 as well; every operator to the left of $T - \lambda_kI$ will then map the result to 0, resulting in the final equivalence. Thus, $(T - \lambda_kI)v' = (T - \lambda_kI)0$ but $v' \neq 0$ for this operator, meaning that it is not injective. But if $T - \lambda_kI$ is not injective, Theorem 5.1 implies that λ_k is an eigenvalue of T, as desired. ∎

- **Matrix** (of $T \in \mathcal{L}(V)$ with respect to the basis $v_1, \dots, v_n$ of V): The n-by-n matrix $\mathcal{M}(T)$ whose entries $A_{j,k}$ are defined by

  $$Tv_k = A_{1,k}v_1 + \cdots + A_{n,k}v_n$$

  - If the basis is not clear from context, then the notation $\mathcal{M}(T, (v_1, \dots, v_n))$ is used.
- "A central goal of linear algebra is to show that given an operator $T \in \mathcal{L}(V)$, there exists a basis of V with respect to which T has a reasonably simple [sparse] matrix" (Axler, 2015, p. 146).

[1]Note that $m \neq n$ necessarily, because a_n may equal 0.

- For example, given just Theorems 3.17 and 5.5, we know that we can choose a basis of V such that the first column of $\mathcal{M}(T)$ for an arbitrary T will have an eigenvalue λ of T in the first row and zeros everywhere else. Specifically, let v_1 be the eigenvector corresponding to λ (guaranteed to exist by Theorem 5.5). Then by Theorem 3.17,

$$\begin{aligned}\mathcal{M}(T)_{\cdot,1} &= \mathcal{M}(Tv_1)\\ &= \mathcal{M}(\lambda v_1)\\ &= \lambda\mathcal{M}(v_1) + 0\mathcal{M}(v_2) + \cdots + 0\mathcal{M}(v_n)\end{aligned}$$

where we extend v_1 into a basis $v_1, \dots, v_n$ of V.

- **Diagonal** (of a square matrix): The entries along the line from the upper left corner to the bottom right corner.
- **Upper-triangular matrix**: A matrix such that all the entries below the diagonal equal 0. *Denoted by*

$$\begin{pmatrix}\lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n\end{pmatrix}$$ [2]

9/22: • Conditions on T that imply that $\mathcal{M}(T)$ is upper triangular.

Theorem 5.6. *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*) *$Tv_j \in \operatorname{span}(v_1, \dots, v_j)$ for each $j = 1, \dots, n$.*

(*c*) *$\operatorname{span}(v_1, \dots, v_j)$ is invariant under T for each $j = 1, \dots, n$.*

Proof. We will first prove that (a) implies (b); the proof in the reverse direction will be symmetric. We will then prove that (b) implies (c), and wrap up by quickly showing that (c) implies (b), and. Let's begin.

Suppose first that the matrix of T with respect to $v_1, \dots, v_n$ is upper triangular. Let $j \in \{1, \dots, n\}$ be arbitrary. By the definition of $\mathcal{M}(T)$, $Tv_j = A_{1,j}v_1 + \cdots + A_{n,j}v_n$. Additionally, since $\mathcal{M}(T)$ is upper triangular, we know that $A_{j+1,j} = \cdots = A_{n,j} = 0$. Therefore, $Tv_j = A_{1,j}v_1 + \cdots + A_{j,j}v_j \in \operatorname{span}(v_1, \dots, v_j)$, as desired. The proof is symmetric in the other direction.

Now suppose that $Tv_j \in \operatorname{span}(v_1, \dots, v_j)$ for each $j = 1, \dots, n$. Fix $j \in \{1, \dots, n\}$. Let $v \in \operatorname{span}(v_1, \dots, v_j)$. It follows since $Tv_i \in \operatorname{span}(v_1, \dots, v_i) \subset \operatorname{span}(v_1, \dots, v_j)$ for all $i = 1, \dots, j$ by hypothesis that

$$\begin{aligned}Tv &= T(a_1v_1 + \cdots + a_jv_j)\\ &= a_1Tv_1 + \cdots + a_jTv_j\\ &\in \operatorname{span}(v_1, \dots, v_j)\end{aligned}$$

as desired. On the other hand, suppose $\operatorname{span}(v_1, \dots, v_j)$ is invariant under T for each $j = 1, \dots, n$. Fix $j \in \{1, \dots, n\}$. Then by the definition of invariance, $v_j \in \operatorname{span}(v_1, \dots, v_j)$ implies that $Tv_j \in \operatorname{span}(v_1, \dots, v_j)$, as desired. ∎

- We now prove that over $\mathbb{C}$ (notably not over $\mathbb{R}$), every operator has an upper-triangular matrix.

Theorem 5.7. *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.*[3]

[2] We often use $*$ to denote matrix entries that we do not know about or that are irrelevant to the questions being discussed.

[3] Note that Axler (2015) gives two proofs, but only one (the shorter one using more advanced topics) is transcribed here.

Proof. We induct on $n = \dim V$. For the base case $n = 1$, the desired result holds since any one-dimensional matrix is upper-triangular by definition. Now suppose inductively that we have proven the result for $n-1$; we wish to prove it for n. By Theorem 5.5, we may let v_1 be an eigenvector of T. Thus, if we let $U = \operatorname{span}(v_1)$, we know that U is an invariant subspace of T with $\dim U = 1$.

Now consider V/U. It follows from the above by Theroem 3.25 that $\dim V/U = n-1$. Consequently, we may apply the inductive hypothesis to learn that $T/U \in \mathcal{L}(V/U)$ has an upper-triangular matrix with respect to some basis $v_2 + U, \dots, v_n + U$ of V/U. It follows by 5.6 that

$$(T/U)(v_j + U) \in \operatorname{span}(v_2 + U, \dots, v_j + U)$$

for each $j = 2, \dots, n$. Thus, since $Tv_1 = \lambda v_1 \in \operatorname{span}(v_1)$ and since the above implies that $Tv_j \in \operatorname{span}(v_2, \dots, v_j) \subset \operatorname{span}(v_1, \dots, v_j)$ for all $j = 2, \dots, n$, we have that $Tv_j \in \operatorname{span}(v_1, \dots, v_j)$ for all $j = 1, \dots, n$. Therefore, by Theorem 5.6, we have that the matrix of T with respect to $v_1, \dots, v_n$ is upper triangular, as desired. ■

- We now prove an easy method for determining invertibility from an upper-triangular matrix.

Theorem 5.8. *Suppose $T \in \mathcal{L}(V)$ has an upper-triangular matrix with respect to some basis of V. Then T is invertible if and only if all the entries on the diagonal of that upper-triangular matrix are nonzero.*

Proof. Let $v_1, \dots, v_n$ be the basis of V with respect to which $\mathcal{M}(T)$ is upper triangular, and let $\lambda_1, \dots, \lambda_n$ be the diagonal entries of said matrix.

Suppose first that T is invertible, and suppose for the sake of contradiction that $\lambda_j = 0$ for some $j = 1, \dots, n$. Then by the definition of $\mathcal{M}(T)$, we have that T maps $\operatorname{span}(v_1, \dots, v_j)$ into $\operatorname{span}(v_1, \dots, v_{j-1})$. Thus, since $\dim \operatorname{span}(v_1, \dots, v_j) = j$ and $\dim \operatorname{span}(v_1, \dots, v_{j-1}) = j - 1$, Theorem 3.7 implies that the restriction of T to $\operatorname{span}(v_1, \dots, v_j)$ is not injective. Consequently, by Theorem 3.4, there exists a nonzero $v \in \operatorname{span}(v_1, \dots, v_j)$ such that $Tv = 0$. But this implies that the original T is not injective, i.e., not invertible (by Theorem 3.13), contradicting our supposition that T is invertible, as desired.

Now suppose that $\lambda_j \neq 0$ for all $j = 1, \dots, n$. To prove that T is invertible, Theorem 3.19 tells us that it will suffice to show that T is surjective. To verify that $\operatorname{range} T = V$, we will demonstrate that $v_1, \dots, v_n \in \operatorname{range} T$, from which it will follow by Theorem 3.5 that $\operatorname{range} T = V$. Let's begin. Since $\mathcal{M}(T)$ is upper triangular, we have that

$$\begin{aligned} Tv_1 &= \lambda_1 v_1 + 0v_2 + \cdots + 0v_n \\ v_1 &= T(v_1/\lambda_1) \in \operatorname{range} T \end{aligned}$$

Similarly, we have that

$$\begin{aligned} Tv_2 &= av_1 + \lambda_2 v_2 + 0v_3 + \cdots + 0v_n \\ v_2 &= T(v_2/\lambda_2) - \frac{a}{\lambda_2} v_1 \in \operatorname{range} T \end{aligned}$$

since $T(v_2/\lambda_2)$ clearly and linear combinations of v_1 are elements of $\operatorname{range} T$ by the previous result. We may analogously prove that $v_3, \dots, v_n \in \operatorname{range} T$, as desired. ■

- We cannot exactly compute the eigenvalues of a linear operator from its matrix, but we can approximate them with powerful numerical methods.
- Lastly, we will show that we can use upper-triangular matrices to determine the eigenvalues of the linear operator it represents.

Theorem 5.9. *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. Let $v_1, \dots, v_n$ be the basis of V with respect to which $\mathcal{M}(T)$ is upper triangular, and let $\lambda_1, \dots, \lambda_n$ be the diagonal entries of said matrix.

Let $\lambda \in \mathbb{F}$. Then $\mathcal{M}(T - \lambda I)$ has diagonal entries $\lambda_1 - \lambda, \dots, \lambda_n - \lambda$. Thus, by Theorem 5.8, $T - \lambda I$ is not invertible iff $\lambda = \lambda_j$ for some $j = 1, \dots, n$. But by Theorem 5.1, it follows that λ is an eigenvalue of T iff $T - \lambda I$ is not invertible iff $\lambda = \lambda_j$ for some $j = 1, \dots, n$, as desired. ∎

5.C Eigenspaces and Diagonal Matrices

- **Diagonal matrix**: A square matrix that is 0 everywhere except possibly along the diagonal.
  - Naturally, every diagonal matrix is upper triangular.
- **Eigenspace** (of $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$): The set of all eigenvectors of T corresponding to λ, along with the zero vector; in other words, the vector space $\text{null}(T - \lambda I)$. *Denoted by* $\boldsymbol{E(\lambda, T)}$.
- We now show that the sum of the eigenspaces is a direct sum.

Theorem 5.10. *Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Suppose also that $\lambda_1, \dots, \lambda_m$ are distinct eigenvalues of T. Then $E(\lambda_1, T) + \cdots + E(\lambda_m, T)$ is a direct sum. Furthermore,* $\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) \leq \dim V$.

Proof. To prove that $E(\lambda_1, T) + \cdots + E(\lambda_m, T)$ is a direct sum, suppose that $u_1 + \cdots + u_m = 0$, where $u_j \in E(\lambda_j, T)$ for all $j = 1, \dots, m$. Naturally, each u_j is an eigenvector corresponding to λ_j or zero. But since Theorem 5.2 implies that $u_1, \dots, u_j$ is linearly independent if any u_j is nonzero, we must have $u_1 = \cdots = u_m = 0$.

Additionally, we have that

$$\begin{aligned}
\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) &= \dim(E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)) \\
&\leq \dim V
\end{aligned}$$

as desired. ∎

- **Diagonalizable** ($T \in \mathcal{L}(V)$): An operator $T \in \mathcal{L}(V)$ such that the operator has a diagonal matrix with respect to some basis of V.
- We now give some conditions equivalent to diagonalizability.

Theorem 5.11. *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) *There exist one-dimensional subspaces $U_1, \dots, U_n$ of V, each invariant under T, such that $V = U_1 \oplus \cdots \oplus U_n$.*

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

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

Proof. We will first briefly justify the equivalence of (a) and (b). We will then show that (b) implies (c) and vice versa. Lastly, we will show that (b) implies (d), (d) implies (e), and (e) implies (b). Let's begin.

Suppose that T is diagonalizable. Then there exists a basis $v_1, \dots, v_n$ of V such that the matrix of T with respect to this basis is diagonal. Let $\lambda_1, \dots, \lambda_n$ be the diagonal entries of this matrix. Then by the definition of $\mathcal{M}(T)$, $Tv_j = \lambda_j v_j$ for all $j = 1, \dots, n$. It follows that each v_j in the basis $v_1, \dots, v_n$ of V is an eigenvector of T, as desired. The proof is symmetric in the reverse direction.

Suppose that $v_1, \dots, v_n$ is a basis of V consisting of eigenvectors of T. Let $U_j = \operatorname{span}(v_j)$ for each $j = 1, \dots, n$. Then each U_j is a one-dimensional subspace of V. Additionally, for any $av_j \in U_j$, we have that $T(av_j) = aTv_j = a\lambda_j v_j \in U_j$, proving that each U_j is invariant under T. Finally, to show that $V = U_1 \oplus \cdots \oplus U_n$, it will suffice to show that each $v \in V$ can be written uniquely as a sum $u_1 + \cdots + u_n$, where $u_j \in U_j$ for each $j = 1, \dots, n$. Let $v \in V$ be arbitrary. Since $v_1, \dots, v_n$ is a basis of V, Theorem 2.5 v can be written uniquely in the form $a_1v_1 + \cdots + a_nv_n$. Let $u_j = a_jv_j$ for each $j = 1, \dots, n$. Then v can be written uniquely as a sum $u_1 + \cdots + u_n$, where $u_j \in U_j$ for each $j = 1, \dots, n$, as desired. The proof in the reverse direction is quite similar.

Suppose that $v_1, \dots, v_n$ is a basis of V consisting of eigenvectors of T. Thus, since each $v \in V$ is a linear combination of eigenvectors of T, $V = E(\lambda_1, T) + \cdots + E(\lambda_m, T)$. Therefore, Theorem 5.10 implies that $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$, as desired.

Suppose $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$. Then naturally $\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T)$, as desired.

Suppose $\dim V = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T)$. Choose a basis of each $E(\lambda_j, T)$. Concatenate all these bases to form a list $v_1, \dots, v_n$ of eigenvectors of T. We now wish to show that $v_1, \dots, v_n$ is a basis of V. To do so, already knowing that $n = \dim V$ by hypothesis, Theorem 2.12 tells us that we need only additionally show that $v_1, \dots, v_n$ is linearly independent. Let $a_1, \dots, a_n \in \mathbb{F}$ be such that $a_1v_1 + \cdots + a_nv_n = 0$. For each $j = 1, \dots, m$, let u_j denote the sum of all the terms a_kv_k such that $v_k \in E(\lambda_j, T)$. Thus, each $u_j \in E(\lambda_j, T)$ and $u_1 + \cdots + u_m = 0$. But since $u_1, \dots, u_m$ are eigenvectors of T corresponding to distinct eigenvalues, Theorem 5.2 implies that $u_1, \dots, u_m$ are linearly independent. This combined with the previous result implies that $u_j = 0$ for each $j = 1, \dots, m$. It follows since each (zero) u_j is a linear combination of a basis of an eigenspace, each $a_j = 0$, as desired. ■

- Note that not every operator is diagonalizable.
- We now prove that if an operator has sufficiently many eigenvalues, it is diagonalizable.

Theorem 5.12. *If $T \in \mathcal{L}(V)$ has $\dim V$ distinct eigenvalues, then T is diagonalizable.*

Proof. Let $m = \dim V$, and let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T. Let $v_1, \dots, v_m$ be the corresponding eigenvectors. Since $v_1, \dots, v_m$ has length equal to the dimension of V and is linearly independent (by Theorem 5.2), Theorem 2.12 implies that $v_1, \dots, v_m$ is a basis of V. Therefore, by Theorem 5.11, T is diagonalizable. ■

- Note that the converse of Theorem 5.12 is not true.

Chapter 6

Inner Product Spaces

6.A Inner Products and Norms

9/30:
- **Norm** (of $x \in \mathbb{R}^n$): The length of x. *Denoted by* $\|x\|$. *Given by*

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$$

- **Dot product** (of $x, y \in \mathbb{R}^n$): The quantity

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

- Properties of the dot product:
  - $x \cdot x = \|x\|^2$.
  - $x \cdot x \geq 0$.
  - $x \cdot x = 0$ iff $x = 0$.
  - Let $y \in \mathbb{R}^n$. Then $T : \mathbb{R}^n \to \mathbb{R}$ defined by $Tx = x \cdot y$ is linear.
  - $x \cdot y = y \cdot x$.
- **Norm** (of $z \in \mathbb{C}^n$): The quantity

$$\|z\| = \sqrt{|z_1|^2 + \cdots + |z_n|^2}$$

  - Note that $\|z\|^2 = z \cdot \bar{z}$.
- **Inner product** (on V): A function that takes each ordered pair $(u, v) \in V \times V$ to a number $\langle u, v \rangle \in \mathbb{F}$ and has the following properties.

  positivity
  $\langle v, v \rangle \geq 0$ for all $v \in V$.

  definiteness
  $\langle v, v \rangle = 0$ iff $v = 0$.

  additivity in first slot
  $\langle u + v, v \rangle = \langle u, w \rangle + \langle v, w \rangle$ for all $u, v, w \in V$.

  homogeneity in first slot
  $\langle \lambda u, v \rangle = \lambda \langle u, v \rangle$ for all $\lambda \in \mathbb{F}$ and all $u, v \in V$.

  conjugate symmetry
  $\langle u, v \rangle = \overline{\langle v, u \rangle}$ for all $u, v \in V$.

- Since every real number equals its complex conjugate, if V is real, we can dispense with the conjugacy condition in the conjugate symmetry condition and just have $\langle u, v\rangle = \langle v, u\rangle$.
- "Although most mathematicians define an inner product as above, many physicists use a definition that requires homogeneity in the second slot instead of the first" (Axler, 2015, p. 166).

- **Euclidean inner product** (on $\mathbb{F}^n$): The function defined by

$$\langle w, z\rangle = w_1\bar{z_1} + \cdots + w_n\bar{z_n}$$

- **Inner product space**: A vector space V along with an inner product on V.
  - When $\mathbb{F}^n$ is referred to as an inner product space, assume that the inner product is the Euclidean inner product unless explicitly stated otherwise.

- Basic properties of an inner product.

Theorem 6.1.

(a) *For each fixed $u \in V$, the function that takes v to $\langle v, u\rangle$ is a linear map from V to $\mathbb{F}$.*

Proof. Let $u \in V$ be arbitrary, and let $T : V \to \mathbb{F}$ be defined by $Tv = \langle v, u\rangle$. Let $v, w \in V$ be arbitrary, and let $\lambda \in \mathbb{F}$. Then

$$\begin{aligned} T(v+w) &= \langle v+w, u\rangle \\ &= \langle v, u\rangle + \langle w, u\rangle \\ &= Tv + Tw \end{aligned} \qquad \begin{aligned} T(\lambda v) &= \langle \lambda v, u\rangle \\ &= \lambda \langle v, u\rangle \\ &= \lambda Tv \end{aligned}$$

as desired. ■

(b) *$\langle 0, u\rangle = 0$ for every $u \in V$.*

Proof. Let $u \in V$ be arbitrary. Since T as defined above is linear, Theorem 3.2 implies that $0 = T(0) = \langle 0, u\rangle$, as desired. ■

(c) *$\langle u, 0\rangle = 0$ for every $u \in V$.*

Proof. Let $u \in V$ be arbitrary. By the conjugate symmetry property and the above, $\langle u, 0\rangle = \overline{\langle 0, u\rangle} = \bar{0} = 0$, as desired. ■

(d) *$\langle u, v+w\rangle = \langle u, v\rangle + \langle u, w\rangle$ for all $u, v, w \in V$.*

Proof. Let $u, v, w \in V$ be arbitrary. Then

$$\begin{aligned} \langle u, v+w\rangle &= \overline{\langle v+w, u\rangle} \\ &= \overline{\langle v, u\rangle + \langle w, u\rangle} \\ &= \overline{\langle v, u\rangle} + \overline{\langle w, u\rangle} && \text{Theorem 4.1} \\ &= \langle u, v\rangle + \langle v, w\rangle \end{aligned}$$

as desired. ■

(e) *$\langle u, \lambda v\rangle = \bar{\lambda}\langle u, v\rangle$ for all $\lambda \in \mathbb{F}$ and $u, v \in V$.*

Proof. Let $u, v \in V$, and let $\lambda \in \mathbb{F}$. Then

$$\begin{aligned} \langle u, \lambda v\rangle &= \overline{\langle \lambda v, u\rangle} \\ &= \overline{\lambda \langle v, u\rangle} \\ &= \bar{\lambda}\overline{\langle v, u\rangle} && \text{Theorem 4.1} \\ &= \bar{\lambda}\langle u, v\rangle \end{aligned}$$

■

- **Norm** (of $v \in V$): The quantity
$$\|v\| = \sqrt{\langle v, v \rangle}$$
- Basic properties of the norm.

  Theorem 6.2. *Suppose $v \in V$. Then*

  (*a*) $\|v\| = 0$ *iff* $v = 0$.

  Proof. Suppose first that $\|v\| = 0$. Then $0 = \sqrt{\langle v, v \rangle} = \langle v, v \rangle$. Thus, $v = 0$. The proof is symmetric in the reverse direction. ■

  (*b*) $\|\lambda v\| = |\lambda|\|v\|$ *for all* $\lambda \in \mathbb{F}$.

  Proof. Let $\lambda \in \mathbb{F}$ be arbitrary. Then
$$\begin{aligned}
\|\lambda v\|^2 &= \langle \lambda v, \lambda v \rangle \\
&= \lambda \bar{\lambda} \langle v, v \rangle \\
&= |\lambda|^2 \|v\|^2 && \text{Theorem 4.1}
\end{aligned}$$
  Taking square roots of the above gives the desired equality.[1] ■

- **Orthogonal** (vectors $u, v \in V$): Two vectors $u, v \in V$ such that $\langle u, v \rangle = 0$.[2]
- If $u, v \in \mathbb{R}^2$ are nonzero, then $\langle u, v \rangle = \|u\|\|v\| \cos \theta$.
  - "Thus, two vectors in $\mathbb{R}^2$ are orthogonal (with respect to the usual Euclidean inner product) if and only if the cosine of the angle between them is 0, which happens if and only if the vectors are perpendicular in the usual sense of plane geometry. Thus, you can think of the word *orthogonal* as a fancy word meaning *perpendicular*." (Axler, 2015, p. 169).
- Orthogonality and zero.

  Theorem 6.3.

  (*a*) 0 *is orthogonal to every vector in* V.

  Proof. Let $u \in V$ be arbitrary. Then Theorem 6.1b implies that $\langle 0, u \rangle = 0$. Thus, u and 0 are orthogonal, as desired. ■

  (*b*) 0 *is the only vector in* V *that is orthogonal to itself.*

  Proof. Let $v \in V$ be such that v is orthogonal to itself. Then $\langle v, v \rangle = 0$. But by the property of definiteness, it follows that $v = 0$, as desired. ■

- The special case where $V = \mathbb{R}^2$ of the following is over 2500 years old, although the following is not the original proof.

  Theorem 6.4 (Pythagorean Theorem)**.** *Suppose $u, v \in V$ are orthogonal. Then*
$$\|u + v\|^2 = \|u\|^2 + \|v\|^2$$
  Proof. We have
$$\begin{aligned}
\|u + v\|^2 &= \langle u + v, u + v \rangle \\
&= (\langle u, u \rangle + \langle u, v \rangle) + (\langle v, u \rangle + \langle v, v \rangle) \\
&= \|u\|^2 + 0 + 0 + \|v\|^2 \\
&= \|u\|^2 + \|v\|^2
\end{aligned}$$
  as desired. ■

[1] Notice this technique: Working with norms squared is usually easier than working directly with norms.
[2] The word *orthogonal* derives from the Greek word *orthogonios*, which means right-angled.

– Note that we can prove the converse of the Pythagorean Theorem in *real* inner product spaces as follows: If $\|u+v\|^2 = \|u\|^2 + \|v\|^2$ for u, v real, then

$$\begin{aligned}
\langle u,u\rangle + \langle v,v\rangle &= \langle u+v, u+v\rangle \\
\langle u,u\rangle + \langle v,v\rangle &= \langle u,u\rangle + \langle v,u\rangle + \langle u,v\rangle + \langle v,v\rangle \\
0 &= \langle u,v\rangle + \langle v,u\rangle \\
&= \langle u,v\rangle + \overline{\langle u,v\rangle} \\
&= 2\operatorname{Re}\langle u,v\rangle \\
&= 2\langle u,v\rangle \\
&= \langle u,v\rangle
\end{aligned}$$

where the first equality holds by the definition of the norm, the second by the additivity of the inner product and Theorem 6.1d, the fifth by Theorem 4.1, and the sixth by the hypothesis that the inner product space is real, as desired.

- Let $u, v \in V$ with $v \neq 0$. We are now equipped to consider how to write u as the sum of v plus a vector w orthogonal to v, as in Figure 6.1 below.

Theorem 6.5. *Suppose $u, v \in V$ with $v \neq 0$. Set $c = \frac{\langle u,v\rangle}{\|v\|^2}$ and $w = u - \frac{\langle u,v\rangle}{\|v\|^2}v$. Then $\langle w, v\rangle = 0$ and $u = cv + w$.*

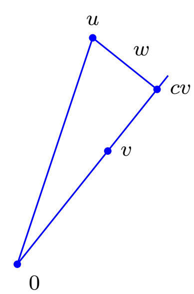

Figure 6.1: An orthogonal decomposition.

Proof. We want to write u in the form $u = cv + w$ where w is orthogonal to v. We know that

$$u = cv + (u - cv)$$

so we need only choose c such that v is orthogonal to $u - cv$. In other words, we want

$$\begin{aligned}
0 &= \langle u - cv, v\rangle \\
&= \langle u,v\rangle + \langle -cv, v\rangle \\
&= \langle u,v\rangle - c\langle v,v\rangle \\
&= \langle u,v\rangle - c\|v\|^2 \\
c &= \frac{\langle u,v\rangle}{\|v\|^2}
\end{aligned}$$

But this gives the values we want for c and w, as desired. ■

- This allows for the proof of a very important result.

Theorem 6.6 (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. We divide into two cases ($v = 0$ and $v \neq 0$). If $v = 0$, then

$$|\langle u, v\rangle| = 0 \leq 0 = \|u\|\sqrt{0} = \|u\|\sqrt{\langle u, v\rangle} = \|u\|\|v\|$$

and we also have that the equality holds since $v = 0 = 0u$, $0 \in \mathbb{F}$. Now let $v \neq 0$. Then by Theorem 6.5,

$$u = \frac{\langle u, v\rangle}{\|v\|^2}v + w$$

where $\langle v, w\rangle = 0$. It follows by the Pythagorean Theorem that

$$\begin{aligned}
\|u\|^2 &= \left\|\frac{\langle u, v\rangle}{\|v\|^2}v\right\|^2 + \|w\|^2 \\
&= \frac{|\langle u, v\rangle|^2}{\|v\|^4}\|v\|^2 + \|w\|^2 && \text{Theorem 6.2b} \\
&= \frac{|\langle u, v\rangle|^2}{\|v\|^2} + \|w\|^2 \\
&\geq \frac{|\langle u, v\rangle|^2}{\|v\|^2}
\end{aligned}$$

Multiplying both sides by $\|v\|^2$ and taking square roots gives the desired inequality.

Also note that the Cauchy-Schwarz inequality is an equality iff the last line is an equality, which happens iff $w = 0$. But $w = 0$ iff u is a scalar multiple of v, as desired. ■

- Note that the Cauchy-Schwarz is known as such because the French mathematician Augustin-Louis Cauchy proved the top inequality below in 1821, and the German mathematician Hermann Schwarz proved the bottom inequality below in 1886; both are special cases of the above.

$$|x_1y_1 + \cdots + x_ny_n|^2 \leq (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2)$$
$$\left|\int_{-1}^1 f(x)g(x)\,\mathrm{d}x\right|^2 \leq \left(\int_{-1}^1 (f(x))^2\,\mathrm{d}x\right)\left(\int_{-1}^1 (g(x))^2\,\mathrm{d}x\right)$$

  - For the top one, we let $x_1, \dots, x_n, y_1, \dots, y_n \in \mathbb{R}$.
  - For the bottom one, we let f, g be continuous real-valued functions on $[-1, 1]$.

- We now prove another important inequality.

Theorem 6.7 (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 multiple of the other.

Proof. We have

$$\begin{aligned}
\|u + v\|^2 &= \langle u + v, u + v\rangle \\
&= \langle u, u\rangle + \langle v, v\rangle + \langle u, v\rangle + \langle v, u\rangle \\
&= \langle u, u\rangle + \langle v, v\rangle + \langle u, v\rangle + \overline{\langle u, v\rangle} \\
&= \|u\|^2 + \|v\|^2 + 2\operatorname{Re}\langle u, v\rangle \\
&\leq \|u\|^2 + \|v\|^2 + 2|\langle u, v\rangle| \\
&\leq \|u\|^2 + \|v\|^2 + 2\|u\|\|v\| && \text{Cauchy-Schwarz Inequality} \\
&= (\|u\| + \|v\|)^2
\end{aligned}$$

Taking square roots of both sides gives the desired inequality.

This inequality is an equality iff $\langle u, v\rangle = \|u\|\|v\|$. Now suppose $u = cv$ where $c \in \mathbb{F}$ is positive. Then

$$\begin{aligned}\langle u, v\rangle &= \langle cv, v\rangle \\ &= c\,\langle v, v\rangle \\ &= c\|v\|^2 \\ &= \|cv\|\|v\| \\ &= \|u\|\|v\|\end{aligned}$$

The proof is the same in the reverse direction. ■

- One last equality.

Theorem 6.8 (Parallelogram Equality)**.** *Suppose $u, v \in V$. Then*

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

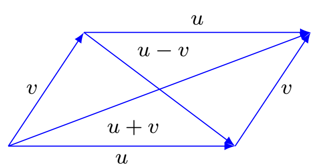

Figure 6.2: The parallelogram equality.

Proof. We have

$$\begin{aligned}\|u+v\|^2 + \|u-v\|^2 &= \langle u+v, u+v\rangle + \langle u-v, u-v\rangle \\ &= \|u\|^2 + \|v\|^2 + \langle u, v\rangle + \langle v, u\rangle + \|u\|^2 + \|v\|^2 - \langle u, v\rangle - \langle v, u\rangle \\ &= 2(\|u\|^2 + \|v\|^2)\end{aligned}$$

as desired. ■

Exercises

10/11: **19** Suppose V is a real inner product space. Prove that

$$\langle u, v\rangle = \frac{\|u+v\|^2 - \|u-v\|^2}{4}$$

for all $u, v \in V$.

Proof. Let $u, v \in V$ be arbitrary. Then

$$\begin{aligned}\frac{\|u+v\|^2 - \|u-v\|^2}{4} &= \frac{\langle u+v, u+v\rangle - \langle u-v, u-v\rangle}{4} \\ &= \frac{(\langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle) - (((\langle u, u\rangle - \langle u, v\rangle) - (\langle v, u\rangle - \langle v, v\rangle))}{4} \\ &= \frac{2\,\langle u, v\rangle + 2\,\langle v, u\rangle}{4} \\ &= \frac{2\,\langle u, v\rangle + 2\,\langle u, v\rangle}{4} \\ &= \langle u, v\rangle\end{aligned}$$

as desired. ■

20 Suppose V is a complex inner product space. Prove that

$$\langle u,v\rangle = \frac{\|u+v\|^2 - \|u-v\|^2 + \|u+iv\|^2 i - \|u-iv\|^2 i}{4}$$

for all $u,v \in V$.

Proof. Let $u,v \in V$ be arbitrary. Then

$$\begin{aligned}
\frac{\|u+v\|^2 - \|u-v\|^2 + \|u+iv\|^2 i - \|u-iv\|^2 i}{4} &= \frac{\|u+v\|^2 - \|u-v\|^2}{4} + i\frac{\|u+iv\|^2 - \|u-iv\|^2}{4} \\
&= \frac{2\langle u,v\rangle + 2\langle v,u\rangle}{4} + i\frac{2\langle u,iv\rangle + 2\langle iv,u\rangle}{4} \\
&= \frac{2\langle u,v\rangle + 2\langle v,u\rangle + 2\langle u,v\rangle + 2\langle -v,u\rangle}{4} \\
&= \frac{4\langle u,v\rangle + 2\langle 0,u\rangle}{4} \\
&= \langle u,v\rangle
\end{aligned}$$

as desired. ■

6.B Orthonormal Bases

10/3:

- **Orthonormal** (list of vectors): A list of vectors such that each vector in the list has norm 1 and is orthogonal to all the other vectors in the list.
  - In other words, a list $e_1,\dots,e_m$ is orthonormal if

$$\langle e_j,e_k\rangle = \begin{cases} 1 & j=k \\ 0 & j\neq k \end{cases}$$

- Orthonormal lists are particularly easy to work with.

Theorem 6.9. *If $e_1,\dots,e_m$ is an orthonormal list of vectors in V, then*

$$\|a_1e_1+\cdots+a_me_m\|^2 = |a_1|^2+\cdots+|a_m|^2$$

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

Proof. We have that

$$\begin{aligned}
\|a_1e_1+\cdots+a_me_m\|^2 &= \|a_1e_1\|^2+\cdots+\|a_me_m\|^2 && \text{Pythagorean Theorem} \\
&= |a_1|^2\|e_1\|^2+\cdots+|a_m|^2\|e_m\|^2 && \text{Theorem 6.2b} \\
&= |a_1|^2+\cdots+|a_m|^2
\end{aligned}$$

as desired. ■

- The next result directly follows from the previous one.

Theorem 6.10. *Every orthonormal list of vectors is linearly independent.*

Proof. Let $e_1,\dots,e_m$ be an orthonormal list of vectors, and let $a_1,\dots,a_m \in \mathbb{F}$ be such that $a_1e_1+\cdots+a_me_m = 0$. Then

$$\begin{aligned}
0 &= \|0\| \\
&= \|a_1e_1+\cdots+a_me_m\| \\
&= |a_1|^2+\cdots+|a_m|^2 && \text{Theorem 6.9}
\end{aligned}$$

But this implies that each $a_j = 0$, as desired. ■

- **Orthonormal basis** (of V): An orthonormal list of vectors in V that is also a basis of V.
- We now prove an easy condition for identifying orthonormal bases.

  Theorem 6.11. *Every orthonormal list of vectors in V with length $\dim V$ is an orthonormal basis of V.*

  Proof. Let $e_1, \dots, e_m$ be an orthonormal list of vectors in V with length $\dim V$. It follows by Theorem 6.10 that $e_1, \dots, e_m$ is linearly independent. Therefore, by Theorem 2.12, $e_1, \dots, e_m$ is a basis of V. ■

10/6:

- For an arbitrary basis $e_1, \dots, e_n$ of V, it can be quite troublesome to compute that $a_1, \dots, a_n \in \mathbb{F}$ that make $a_1e_1 + \cdots + a_ne_n = v$ for an arbitrary $v \in V$.
- However, it is quite simple for an orthonormal basis:

  Theorem 6.12. *Suppose $e_1, \dots, e_n$ is an orthonormal basis of V and $v \in V$. Then*

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n$$

  and

$$\|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2$$

  Proof. Because $e_1, \dots, e_n$ is a basis of V, we have that

$$v = a_1e_1 + \cdots + a_ne_n$$

  for some $a_1, \dots, a_n \in \mathbb{F}$. It follows by taking inner products with an arbitrary e_j that

$$\begin{aligned}
\langle v, e_j \rangle &= \langle a_1e_1 + \cdots + a_ne_n, e_j \rangle \\
&= a_1 \langle e_1, e_j \rangle + \cdots + a_n \langle e_n, e_j \rangle \\
&= a_1 \cdot 0 + \cdots + a_{j-1} \cdot 0 + a_j \cdot 1 + a_{j+1} \cdot 0 + \cdots + a_n \cdot 0 \\
&= a_j
\end{aligned}$$

  for each $j = 1, \dots, n$.

  The second equation follows from the first equation by Theorem 6.9. ■

- The following algorithm[3] gives a method for turning a linearly independent list into an orthonormal list with the same span as the original list.

  Theorem 6.13 (Gram-Schmidt Procedure)**.** *Suppose $v_1, \dots, v_m$ is a linearly independent list of vectors in V. Let $e_1 = v_1/\|v_1\|$. For $j = 2, \dots, m$, define e_j inductively by*

$$e_j = \frac{v_j - \langle v_j, e_1 \rangle e_1 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1}}{\|v_j - \langle v_j, e_1 \rangle e_1 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1}\|}$$

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

$$\operatorname{span}(v_1, \dots, v_j) = \operatorname{span}(e_1, \dots, e_j)$$

  for $j = 1, \dots, m$.

  Proof. We induct on j. For the base case $j = 1$, we have that e_1 is trivially orthogonal to all other vectors in the list and that

$$\|e_1\| = \left\| \frac{v_1}{\|v_1\|} \right\| = \frac{1}{\|v_1\|} \|v_1\| = 1$$

[3]Danish mathematician Jørgen Gram and German mathematician Erhard Schmidt popularized this algorithm.

proving that e_1 is an orthonormal list of vectors in V. Additionally, since e_1 is a scalar multiple of v_1, we naturally also have that

$$\operatorname{span}(v_1) = \operatorname{span}(e_1)$$

Now suppose inductively that we have proven that $e_1, \dots, e_j$ is an orthonormal list of vectors such that $\operatorname{span}(v_1, \dots, v_j) = \operatorname{span}(e_1, \dots, e_j)$. We now wish to prove the claim for e_{j+1}. To do so, it will suffice to show that e_{j+1} is well defined, that $e_1, \dots, e_{j+1}$ is an orthonormal list of vectors, and that $\operatorname{span}(v_1, \dots, v_{j+1}) = \operatorname{span}(e_1, \dots, e_{j+1})$. Let's begin.

To confirm that the e_{j+1} is well defined, we must confirm that the norm in the bottom is nonzero. But since $v_1, \dots, v_m$ is linearly independent by hypothesis, we have that $v_{j+1} \notin \operatorname{span}(v_1, \dots, v_j) = \operatorname{span}(e_1, \dots, e_j)$, meaning that we must have

$$\begin{aligned} v_{j+1} &\neq \langle v_{j+1}, e_1 \rangle e_1 + \cdots + \langle v_{j+1}, e_j \rangle e_j \\ v_{j+1} - \langle v_{j+1}, e_1 \rangle e_1 - \cdots - \langle v_{j+1}, e_j \rangle e_j &\neq 0 \\ \| v_{j+1} - \langle v_{j+1}, e_1 \rangle e_1 - \cdots - \langle v_{j+1}, e_j \rangle e_j \| &\neq 0 && \text{Theorem 6.2a} \end{aligned}$$

as desired.

To confirm that $e_1, \dots, e_{j+1}$ is an orthonormal list of vectors, it will suffice to show that e_{j+1} has norm 1 (the inductive hypothesis guarantees that $e_1, \dots, e_j$ have norm 1) and that each vector in $e_1, \dots, e_j$ is orthogonal to e_{j+1} (again, we already know by the inductive hypothesis that $e_1, \dots, e_j$ is mutually orthogonal). We can address the first part with an argument symmetric to that used in the base case (indeed, it is easily seen that any vector divided by its norm has norm 1). As to the second part of the argument, we have for any $1 \leq k < j+1$ that

$$\begin{aligned} \langle e_{j+1}, e_k \rangle &= \left\langle \frac{v_{j+1} - \langle v_{j+1}, e_1 \rangle e_1 - \cdots - \langle v_{j+1}, e_j \rangle e_j}{\| v_{j+1} - \langle v_{j+1}, e_1 \rangle e_1 - \cdots - \langle v_{j+1}, e_j \rangle e_j \|}, e_k \right\rangle \\ &= \frac{1}{\| v_{j+1} - \langle v_{j+1}, e_1 \rangle e_1 - \cdots - \langle v_{j+1}, e_j \rangle e_j \|} (\langle v_{j+1}, e_k \rangle - \langle v_{j+1}, e_1 \rangle \langle e_1, e_k \rangle - \cdots - \langle v_{j+1}, e_j \rangle \langle e_j, e_k \rangle) \\ &= \frac{1}{\| v_{j+1} - \langle v_{j+1}, e_1 \rangle e_1 - \cdots - \langle v_{j+1}, e_j \rangle e_j \|} (\langle v_{j+1}, e_k \rangle - \langle v_{j+1}, e_k \rangle \cdot 1) \\ &= 0 \end{aligned}$$

as desired.

To confirm that $\operatorname{span}(v_1, \dots, v_{j+1}) = \operatorname{span}(e_1, \dots, e_{j+1})$, Exercise 2.C.1 tells us that it will suffice to show that $\operatorname{span}(v_1, \dots, v_{j+1}) \subset \operatorname{span}(e_1, \dots, e_{j+1})$ and that $\dim \operatorname{span}(v_1, \dots, v_{j+1}) = \dim \operatorname{span}(e_1, \dots, e_{j+1})$. Since $\operatorname{span}(v_1, \dots, v_j) = \operatorname{span}(e_1, \dots, e_j)$ by hypothesis, we have that $v_1, \dots, v_j \in \operatorname{span}(e_1, \dots, e_j) \subset \operatorname{span}(e_1, \dots, e_{j+1})$. Additionally, by the definition of e_{j+1}, we have that $v_{j+1} \in \operatorname{span}(e_1, \dots, e_{j+1})$. Therefore, we have that

$$\operatorname{span}(v_1, \dots, v_{j+1}) \subset \operatorname{span}(e_1, \dots, e_{j+1})$$

Additionally, since $v_1, \dots, v_{j+1}$ is linearly independent by hypothesis, and $e_1, \dots, e_{j+1}$ is linearly independent by Theorem 6.10, we have that both subspaces have dimension $j+1$, as desired. ■

- Existence of an orthonormal basis.

Theorem 6.14. *Every finite-dimensional inner product space has an orthonormal basis.*

Proof. Let V be a finite-dimensional inner product space. Choose a basis of V. Applying the Gram-Schmidt Procedure to this (by definition linearly independent) basis yields an orthonormal list of vectors in V of length $\dim V$. Therefore, by Theorem 6.11, the orthonormal list is an orthonormal basis of V. ■

- Extending an orthonormal list to an orthonormal basis.

Theorem 6.15. *Suppose V is finite-dimensional. Then every orthonormal list of vectors in V can be extended to an orthonormal basis of V.*

Proof. Suppose $e_1, \dots, e_m$ is an orthonormal list of vectors in V. Then by Theorem 6.10, $e_1, \dots, e_m$ is linearly independent. Hence this list can be extended to a basis $e_1, \dots, e_m, v_1, \dots, v_n$ of V by Theorem 2.8. Applying the Gram-Schmidt Procedure to this basis yields an orthonormal list $e_1, \dots, e_m, f_1, \dots, f_n$ of vectors in V. Note that the Gram-Schmidt Procedure doesn't change the first m vectors of the list: If $1 \leq j \leq m$, then

$$\begin{aligned} e_j' &= \frac{e_j - \langle e_j, e_1 \rangle e_1 - \cdots - \langle e_j, e_{j-1} \rangle e_{j-1}}{\|e_j - \langle e_j, e_1 \rangle e_1 - \cdots - \langle e_j, e_{j-1} \rangle e_{j-1}\|} \\ &= \frac{e_j}{\|e_j\|} \\ &= e_j \end{aligned}$$

Lastly, by Theorem 6.11, we know that $e_1, \dots, e_m, f_1, \dots, f_n$ is a basis of V. ■

- The next result builds on Theorem 5.7.

Theorem 6.16. *Suppose $T \in \mathcal{L}(V)$. If T has an upper-triangular matrix with respect to some basis of V, then T has an upper-triangular matrix with respect to some orthonormal basis of V.*

Proof. Suppose T has an upper-triangular matrix with respect to the basis $v_1, \dots, v_n$ of V. Then by Theorem 5.6, $\operatorname{span}(v_1, \dots, v_j)$ is invariant under T for each $j = 1, \dots, n$. Applying the Gram-Schmidt Procedure to $v_1, \dots, v_n$, we obtain an orthonormal basis $e_1, \dots, e_n$ of V with the property that

$$\operatorname{span}(e_1, \dots, e_j) = \operatorname{span}(v_1, \dots, v_j)$$

for each $j = 1, \dots, n$. Combining the last two results, we have that $\operatorname{span}(e_1, \dots, e_j)$ is invariant under T for each $j = 1, \dots, n$. Therefore, by Theorem 5.6, T has an upper-triangular matrix with respect to the orthonormal basis $e_1, \dots, e_n$. ■

- We now prove an important application of the above.

Theorem 6.17 (Schur's Theorem[4])**.** *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 orthonormal basis of V.*

Proof. By Theorem 5.7, T has an upper-triangular matrix with respect to some basis of V. Therefore, by Theorem 6.16, T has an upper-triangular matrix with respect to some orthonormal basis of V, as desired. ■

- We now show that every linear functional is equivalent to some inner product.

Theorem 6.18 (Riesz Representation Theorem[5])**.** *Suppose V is finite-dimensional and φ is a linear functional on V. Then there is a unique vector $u \in V$ such that*

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

for every $v \in V$.

[4]German mathematician Issai Schur published the first proof of this result in 1909.

[5]This result is named for Hungarian mathematician Frigyes Riesz, who proved several results in the early twentieth century very similar to this one.

Proof. We first find such a vector u. Let $e_1, \dots, e_n$ be an orthonormal basis of V. Then

$$\begin{aligned}
\varphi(v) &= \varphi(\langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n) && \text{Theorem 6.12}\\
&= \langle v, e_1\rangle \varphi(e_1) + \cdots + \langle v, e_n\rangle \varphi(e_n)\\
&= \left\langle v, \overline{\varphi(e_1)}e_1 + \cdots + \overline{\varphi(e_n)}e_n \right\rangle && \text{Theorem 6.1e}
\end{aligned}$$

so we may take

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

Now suppose for the sake of contradiction that $u_1, u_2 \in V$ are such that

$$\varphi(v) = \langle v, u_1\rangle = \langle v, u_2\rangle$$

for every $v \in V$. Then

$$0 = \langle v, u_1\rangle - \langle v, u_2\rangle = \langle v, u_1 - u_2\rangle$$

for every $v \in V$. In particular, if we let $v = u_1 - u_2$, we have that $\langle u_1 - u_2, u_1 - u_2\rangle = 0$. Therefore, by the definiteness of the inner product, $u_1 - u_2 = 0$, i.e., $u_1 = u_2$, a contradiction. ■

- Note that despite the fact that the expression for u derived in the proof of the Riesz Representation Theorem seems to depend on both φ and the chosen orthonormal basis, the statement of the theorem shows that the value of u depends only on φ, and is in truth the same for any orthonormal basis.

6.C Orthogonal Complements and Minimization Problems

- **Orthogonal complement** (of $U \subset V$): The set of all vectors in V that are orthogonal to every vector in U. *Denoted by* $\boldsymbol{U^\perp}$. *Given by*

$$U^\perp = \{v \in V : \langle v, u\rangle = 0 \ \forall \ u \in U\}$$

- Basic properties of the orthogonal complement.

Theorem 6.19.

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

Proof. Theorem 6.1b tells us that $\langle 0, u\rangle = 0$ for every $u \in U$; thus, $0 \in U^\perp$, as desired.
Let $v, w \in U^\perp$. Then

$$\langle v + w, u\rangle = \langle v, u\rangle + \langle w, u\rangle = 0 + 0 = 0$$

for all $u \in U$. Thus, $v + w \in U^\perp$, as desired.
Let $v \in U^\perp$ and $\lambda \in \mathbb{F}$. Then

$$\langle \lambda v, u\rangle = \lambda \langle v, u\rangle = \lambda \cdot 0 = 0$$

for all $u \in U$. Thus, $\lambda v \in U^\perp$, as desired. ■

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

Proof. Let $v \in V$ be arbitrary. Then since $\langle v, 0\rangle = 0$ by Theorem 6.1c, we have that $v \in \{0\}^\perp$. Thus, $\{0\}^\perp = V$. ■

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

Proof. Suppose $v \in V^\perp$. Then $\langle v, u\rangle = 0$ for all $u \in V$. In particular, $\langle v, v\rangle = 0$. But this implies that $v = 0$ by the definiteness of the inner product. Therefore, $V^\perp = \{0\}$. ■

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

Proof. Let $v \in U \cap U^\perp$. Then $\langle v, v \rangle = 0$. Therefore, $U \cap U^\perp \subset \{0\}$. ■

(e) *If U and W are subsets of V and $U \subset W$, then $W^\perp \subset U^\perp$.*

Proof. Let $v \in W^\perp$. Then $\langle v, u \rangle = 0$ for all $u \in W$. But since $U \subset W$, this implies that $\langle v, u \rangle = 0$ for all $u \in U$. Thus, $v \in U^\perp$. Therefore, $W^\perp \subset U^\perp$. ■

- We can decompose V into the direct sum of any subspace U and another subspace.

Theorem 6.20. *Suppose U is a finite-dimensional subspace of V. Then*

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

Proof. First, we will show that any $v \in V$ can be written as the sum of an element of U and an element of $U^\perp$. Then, we will show that this decomposition is unique. Let's begin.

Let $v \in V$ be arbitrary, and let $e_1, \dots, e_m$ be an orthonormal basis of U. Then we may let

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

Since $e_1, \dots, e_m$ is a basis of U, $u \in U$. Additionally, we have that

$$\begin{aligned}
\langle w, e_j \rangle &= \langle v - \langle v, e_1 \rangle e_1 - \cdots - \langle v, e_m \rangle e_m, e_j \rangle \\
&= \langle v, e_j \rangle - \langle v, e_1 \rangle \langle e_1, e_j \rangle - \cdots - \langle v, e_m \rangle \langle e_m, e_j \rangle \\
&= \langle v, e_j \rangle - \langle v, e_j \rangle \cdot 1 \\
&= 0
\end{aligned}$$

for all e_j. Thus, w is orthogonal to every vector in $\text{span}(e_1, \dots, e_m) = U$. It follows that $w \in U^\perp$. Therefore, $v = u + w$ for $u \in U$ and $w \in U^\perp$, as desired.

This combined with the fact that $U \cap U^\perp = \{0\}$ (by Theorem 6.19b) implies that $V = U \oplus U^\perp$, as desired. ■

- Dimension of the orthogonal complement.

Theorem 6.21. *Suppose V is finite-dimensional and U is a subspace of V. Then*

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

Proof. By Theorem 6.20, we have that $V = U \oplus U^\perp$. Therefore, by Theorem 3.22,

$$\begin{aligned}
\dim V &= \dim U + \dim U^\perp \\
\dim U^\perp &= \dim V - \dim U
\end{aligned}$$

as desired. ■

- We now have an important corollary to Theorem 6.20.

Theorem 6.22. *Suppose U is a finite-dimensional subspace of V. Then*

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

Proof. Suppose first that $u \in U$. Then $\langle u, v \rangle = 0$ for all $v \in U^\perp$ by the definition of $U^\perp$. But this also means that u is orthogonal to every vector in $U^\perp$, i.e., that $u \in (U^\perp)^\perp$. Thus $U \subset (U^\perp)^\perp$.

Now suppose that $v \in (U^\perp)^\perp$. By Theorem 6.20, we have that $v = u + w$, where $u \in U$ and $w \in U^\perp$. It follows that $v - u = w \in U^\perp$ and, since $v \in (U^\perp)^\perp$ by hypothesis and $u \in (U^\perp)^\perp$ by the first part of the proof, that $v - u \in (U^\perp)^\perp$. Therefore, $v - u \in U^\perp \cap (U^\perp)^\perp \subset \{0\}$, so $v - u = 0$ or $v = u \in U$, as desired. ■

- **Orthogonal projection** (of V onto U): The operator $P_U \in \mathcal{L}(V)$ defined by $P_U v = u$ iff $v = u + w$ where $u \in U$ and $w \in U^\perp$.
  - Since Theorem 6.20 implies that each $v \in V$ can be written uniquely in the form $u + w$ where $u \in U$ and $w \in U^\perp$ for any subspace U of V, $P_U v$ is well defined in general.
- Properties of the orthogonal projection.

Theorem 6.23. *Suppose U is a finite-dimensional subspace of V and $v \in V$. Then*

(a) $P_U \in \mathcal{L}(V)$.

Proof. To prove that P_U is a linear map on V, suppose $v_1, v_2 \in V$ and $a_1, a_2 \in \mathbb{F}$. Let

$$v_1 = u_1 + w_1 \qquad\qquad v_2 = u_2 + w_2$$

where $u_1, u_2 \in U$ and $w_1, w_2 \in U^\perp$. Then

$$\begin{aligned} P_U(a_1v_1 + a_2v_2) &= P_U(a_1u_1 + a_2u_2 + a_1w_1 + a_2w_2) \\ &= a_1u_1 + a_2u_2 \\ &= a_1P_Uv_1 + a_2P_Uv_2 \end{aligned}$$

as desired. ∎

(b) $P_U u = u$ *for every* $u \in U$.

Proof. Let $u \in U$ be arbitrary. Then $u = u + 0$ where $u \in U$ and $0 \in U^\perp$. Thus, $P_U u = u$. ∎

(c) $P_U w = 0$ *for every* $w \in U^\perp$.

Proof. Let $w \in U^\perp$. Then $w = 0 + w$ where $0 \in U$ and $w \in U^\perp$. Thus, $P_U w = 0$. ∎

(d) $\operatorname{range} P_U = U$.

Proof. By the definition of P_U, $\operatorname{range} P_U \subset U$. Additionally, by Theorem 6.23b, $U \subset \operatorname{range} P_U$. Thus, $\operatorname{range} P_U = U$, as desired. ∎

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

Proof. Theorem 6.23c implies that $U^\perp \subset P_U$. In the other direction, we have that if $v \in P_U$, then $P_U v = 0$. Thus, we must have $v = 0 + v$, where $0 \in U$ and $v \in U^\perp$. Consequently, $P_U \subset U^\perp$. Therefore, $P_U = U^\perp$, as desired. ∎

(f) $v - P_U v \in U^\perp$.

Proof. If $v = u + w$ with $u \in U$ and $w \in U^\perp$, then $v - P_U v = v - u = w \in U^\perp$, as desired. ∎

(g) $P_U^2 = P_U$.

Proof. If $v - u + w$ with $u \in U$ and $w \in U^\perp$, then $(P_U^2)v = P_U(P_U v) = P_U u = u = P_U v$, as desired. ∎

(h) $\|P_U v\| \leq \|v\|$.

Proof. If $v = u + w$ with $u \in U$ and $w \in U^\perp$, then

$$\begin{aligned} \|P_U v\| &= \|u\|^2 \\ &\leq \|u\|^2 + \|w\|^2 \\ &= \|v\|^2 && \text{Pythagorean Theorem} \end{aligned}$$

Taking square roots gives the desired inequality. ∎

(*i*) *For every orthonormal basis* $e_1, \ldots, e_m$ *of* U,

$$P_U v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m$$

Proof. If $v = u + w$ with $u \in U$ and $w \in U^{\perp}$, then

$$\begin{aligned} P_U v &= u \\ &= \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m && \text{Theorem 6.12} \end{aligned}$$

as desired. ■

- Minimizing the distance to a subspace.

Theorem 6.24. *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

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

Taking square roots of the above give the desired inequality.

The inequality is an equality if and only if $\|P_U v - u\|^2 = 0$. But this happens if and only if $u = P_U v$, as desired. ■

- The above theorem is often combined with Theorem 6.23i to compute explicit solutions to minimization problems.

Chapter 7

Operators on Inner Product Spaces

7.A Self-Adjoints and Normal Operators

10/7:

- **Adjoint** (of $T \in \mathcal{L}(V, W)$): The function $T^* : W \to V$ that satisfies

$$\langle Tv, w \rangle = \langle v, T^* w \rangle$$

for all $v \in V$ and $w \in W$[1].

  - Calculating T^*w: Consider the linear functional $\varphi : V \to \mathbb{F}$ defined by $\varphi(v) = \langle Tv, w \rangle$ for all $v \in V$. By the Riesz Representation Theorem, there exists a unique vector $T^*w \in V$ such that $\varphi(v) = \langle v, T^*w \rangle$ for all $v \in V$. This vector in V will guarantee that $\langle Tv, w \rangle = \varphi(v) = \langle v, T^*w \rangle$ for all $v \in V$, and we can find vectors $T^*w \in V$ for all $w \in W$.

- The adjoint is a linear map.

Theorem 7.1. *If $T \in \mathcal{L}(V, W)$, then $T^* \in \mathcal{L}(W, V)$.*

Proof. Let $T \in \mathcal{L}(V, W)$, let $w_1, w_2 \in W$, and let $\lambda \in \mathbb{F}$. By the definition of T^*, we have that for any $v \in V$,

$$\begin{aligned}
\langle v, T^*(w_1 + w_2) \rangle &= \langle Tv, w_1 + w_2 \rangle \\
&= \langle Tv, w_1 \rangle + \langle Tv, w_2 \rangle \\
&= \langle v, T^*w_1 \rangle + \langle v, T^*w_2 \rangle \\
&= \langle v, T^*w_1 + T^*w_2 \rangle
\end{aligned}
\qquad
\begin{aligned}
\langle v, T^*(\lambda w_1) \rangle &= \langle Tv, \lambda w_1 \rangle \\
&= \bar{\lambda} \langle Tv, w_1 \rangle \\
&= \bar{\lambda} \langle v, T^*w_1 \rangle \\
&= \langle v, \lambda T^*w_1 \rangle
\end{aligned}$$

Thus, by the definition of T^*,

$$T^*(w_1 + w_2) = T^*w_1 + T^*w_2 \qquad T^*(\lambda w_1) = \lambda T^* w$$

so T^* is a linear map, as desired. ∎

- Properties of the adjoint.

Theorem 7.2.

(a) $(S + T)^* = S^* + T^*$ *for all* $S < T \in \mathcal{L}(V, W)$.

[1] Note that the word adjoint has another, unrelated meaning in algebra. Fortunately, this other meaning will not be covered in Axler (2015).

Proof. Suppose $S, T \in \mathcal{L}(V, W)$. If $v \in V$ and $w \in W$, then

$$\begin{aligned}\langle v, (S+T)^* w\rangle &= \langle (S+T)v, w\rangle \\ &= \langle Sv, w\rangle + \langle Tv, w\rangle \\ &= \langle v, S^*w\rangle + \langle v, T^*w\rangle \\ &= \langle v, S^*w + T^*w\rangle\end{aligned}$$

Thus, $(S+T)^*w = S^*w + T^*w$, as desired. ∎

(*b*) $(\lambda T)^* = \bar{\lambda} T^*$ *for all* $\lambda \in \mathbb{F}$ *and* $T \in \mathcal{L}(V, W)$.

Proof. Suppose $T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbb{F}$. If $v \in V$ and $w \in W$, then

$$\begin{aligned}\langle v, (\lambda T)^* w\rangle &= \langle \lambda T v, w\rangle \\ &= \lambda \langle Tv, w\rangle \\ &= \lambda \langle v, T^*w\rangle \\ &= \left\langle v, \bar{\lambda} T^* w\right\rangle\end{aligned}$$

Thus, $(\lambda T)^* w = \bar{\lambda} T^* w$, as desired. ∎

(*c*) $(T^*)^* = T$ *for all* $T \in \mathcal{L}(V, W)$.

Proof. Suppose $T \in \mathcal{L}(V, W)$. If $v \in V$ and $w \in W$, then

$$\begin{aligned}\langle w, (T^*)^* v\rangle &= \langle T^* w, v\rangle \\ &= \overline{\langle v, T^* w\rangle} \\ &= \overline{\langle Tv, w\rangle} \\ &= \langle w, Tv\rangle\end{aligned}$$

Thus, $(T^*)^* v = Tv$, as desired. ∎

(*d*) $I^* = I$, *where* I *is the identity operator on* V.

Proof. If $v, u \in V$, then

$$\langle v, I^* u\rangle = \langle Iv, u\rangle = \langle v, Iu\rangle$$

Thus, $I^* u = Iu$, as desired. ∎

(*e*) $(ST)^* = T^* S^*$ *for all* $T \in \mathcal{L}(V, W)$ *and* $S \in \mathcal{L}(W, U)$. *Here* U *is an inner product space over* $\mathbb{F}$.

Proof. Suppose $T \in \mathcal{L}(V, W)$ and $S \in \mathcal{L}(W, U)$. If $v \in V$ and $u \in U$, then

$$\begin{aligned}\langle v, (ST)^* u\rangle &= \langle STv, u\rangle \\ &= \langle Tv, S^* u\rangle \\ &= \langle v, T^* S^* u\rangle\end{aligned}$$

Thus, $(ST)^* u = T^* S^* u$, as desired. ∎

- Null space and range of T^*.

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

(*a*) $\operatorname{null} T^* = (\operatorname{range} T)^\perp$.

Proof. Let $w \in W$ be an arbitrary element of $\operatorname{null} T^*$. Then $T^* w = 0$ by definition. It follows by Theorem 6.1c that $\langle v, T^* w\rangle = 0$ for all $v \in V$. Thus, by the definition of the adjoint, $\langle Tv, w\rangle = 0$ for all $v \in V$. But this implies that w is orthogonal to every vector in $\operatorname{range} T$ (i.e., the set of all Tv), meaning that $w \in (\operatorname{range} T)^\perp$.
The proof is symmetric in the other direction. ∎

(*b*) $\operatorname{range} T^* = (\operatorname{null} T)^\perp$.

Proof. We have that

$$\begin{aligned} \operatorname{range} T^* &= ((\operatorname{range} T^*)^\perp)^\perp && \text{Theorem 6.22} \\ &= (\operatorname{null}(T^*)^*)^\perp && \text{Theorem 7.3a} \\ &= (\operatorname{null} T)^\perp && \text{Theorem 7.2c} \end{aligned}$$

as desired. ■

(*c*) $\operatorname{null} T = (\operatorname{range} T^*)^\perp$.

Proof. We have that

$$\begin{aligned} \operatorname{null} T &= \operatorname{null}(T^*)^* && \text{Theorem 7.2c} \\ &= (\operatorname{range} T^*)^\perp && \text{Theorem 7.3a} \end{aligned}$$

as desired. ■

(*d*) $\operatorname{range} T = (\operatorname{null} T^*)^\perp$.

Proof. We have that

$$\begin{aligned} \operatorname{range} T &= ((\operatorname{range} T)^\perp)^\perp && \text{Theorem 6.22} \\ &= (\operatorname{null} T^*)^\perp && \text{Theorem 7.3a} \end{aligned}$$

as desired. ■

- **Conjugate transpose** (of an m-by-n matrix): The n-by-m matrix obtained by interchanging the rows and columns and then taking the complex conjugate of each entry.
  - "If $\mathbb{F} = \mathbb{R}$, then the conjugate transpose of a matrix is the same as its transpose" (Axler, 2015, p. 207).
- The next result shows how to compute the matrix of T^* from the matrix of T. Note, however, that if $\mathcal{M}(T)$ is with respect to nonorthonormal bases, $\mathcal{M}(T^*)$ does not necessarily equal the conjugate transpose of $\mathcal{M}(T)$.

Theorem 7.4. *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*

$$\mathcal{M}(T^*, (f_1, \dots, f_m), (e_1, \dots, e_n))$$

is the conjugate transpose of

$$\mathcal{M}(T, (e_1, \dots, e_n), (f_1, \dots, f_m))$$

Proof. Recall that the k^{th} column of $\mathcal{M}(T)$ is given by writing Te_k as a linear combination of the f_j's. Since $f_1, \dots, f_m$ is an orthonormal basis of W, Theorem 6.12 implies that

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

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

$$T^* f_k = \langle T^* f_k, e_1 \rangle e_1 + \cdots + \langle T^* f_k, e_n \rangle e_n$$

we have that the entry in row j column k of $\mathcal{M}(T^*)$ is

$$\begin{aligned} \langle T^* f_k, e_j \rangle &= \langle f_k, Te_j \rangle \\ &= \overline{\langle Te_j, f_k \rangle} \end{aligned}$$

Therefore, the entry in row k column j of $\mathcal{M}(T^*)$ is the complex conjugate of the entry in row j column k of $\mathcal{M}(T)$, as desired. ■

- **Self-adjoint** (operator $T \in \mathcal{L}(V)$): An operator T such that $T = T^*$. *Also known as* **Hermitian**.
  - In other words, $T \in \mathcal{L}(V)$ is self-adjoint if and only if

    $$\langle Tv, w\rangle = \langle v, Tw\rangle$$

    for all $v, w \in V$.
- The sum of two self-adjoint operators is self-adjoint, and the product of a real scalar and a self-adjoint operator is self-adjoint.
- Note the analogy between self-adjoint operator and complex numbers: A complex number z is real iff $z = \bar{z}$, and thus a self-adjoint operator ($T = T^*$) is analogous to a real number.
- Eigenvalues of self-adjoint operators.

  Theorem 7.5. *Every eigenvalue of a self-adjoint operator is real.*

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

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

  so $\lambda = \bar{\lambda}$, which implies that λ is real, as desired. ■

- The next result is false for real inner product spaces (consider a rotation matrix), but true for complex ones.

  Theorem 7.6. *Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$. Suppose $\langle Tv, v\rangle = 0$ for all $v \in V$. Then $T = 0$.*

  Proof. Let $u \in V$ be arbitrary. By inner product algebra, we have that

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

  for all $w \in V$. Since each term on the right-hand side of the above equation is of the form $\langle Tv, v\rangle$ and we know by hypothesis that $\langle Tv, v\rangle = 0$ for all $v \in V$, we have that $\langle Tu, w\rangle = 0$ for all $w \in V$. In particular, if we let $w = Tu$, we learn that $\langle Tu, Tu\rangle = 0$, which implies that $Tu = 0$. But this implies that $Tu = 0$ for all $u \in V$, i.e., that $T = 0$. ■

- The next result provides another example of how self-adjoint operators behave like real numbers, and is also false for real inner product spaces (consider a operator on such a space that is not self-adjoint).

  Theorem 7.7. *Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$. Then T is self-adjoint if and only if $\langle Tv, v\rangle \in \mathbb{R}$ for every $v \in V$.*

  Proof. Suppose first that T is self-adjoint. Let $v \in V$ be arbitrary. Then

  $$\langle Tv, v\rangle - \overline{\langle Tv, v\rangle} = \langle Tv, v\rangle - \langle v, Tv\rangle = \langle Tv, v\rangle - \langle T^*v, v\rangle = \langle (T - T^*)v, v\rangle = \langle 0v, v\rangle = 0$$

  so $\langle Tv, v\rangle = \overline{\langle Tv, v\rangle}$. Therefore, $\langle Tv, v\rangle \in \mathbb{R}$, as desired.

  Now suppose that $\langle Tv, v\rangle \in \mathbb{R}$ for every $v \in V$. Let $v \in V$ be arbitrary. Then

  $$\langle (T - T^*)v, v\rangle = \langle Tv, v\rangle - \langle T^*v, v\rangle = \langle Tv, v\rangle - \langle v, Tv\rangle = \langle Tv, v\rangle - \overline{\langle Tv, v\rangle} = 0$$

  Therefore, by Theorem 7.6, $T - T^* = 0$, or $T = T^*$, as desired. ■

- We now show that on complex *or* real vector spaces, self-adjoint operators that satisfy $\langle Tv, v\rangle = 0$ *must* be the zero operator.

  Theorem 7.8. *Suppose T is a self-adjoint operator on V such that*

  $$\langle Tv, v\rangle = 0$$

  for all $v \in V$. Then $T = 0$.

  Proof. We divide into two cases. If V is complex, invoke Theorem 7.6. If V is real, we continue.

  Let $u \in V$ be arbitrary. By inner product algebra, we have that

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

  By a symmetric argument to that used in the later part of the proof of Theorem 7.6, we can confirm that $T = 0$. ■

- **Normal** (operator): An operator that commutes with its adjoint.
  - In other words, $T \in \mathcal{L}(V)$ is normal if
    $$TT^* = T^*T$$
- Every self-adjoint operator is normal.
- We now characterize normal operators.

  Theorem 7.9. *An operator is normal if and only if*

  $$\|Tv\| = \|T^*v\|$$

  for all $v \in V$.

  Proof. Let $T \in \mathcal{L}(V)$.

  Suppose first that T is normal. Then $T^*T - TT^* = 0$. Thus, by Theorem 6.1b, $\langle (T^*T - TT^*)v, v\rangle = 0$ for all $v \in V$. It follows that

  $$\begin{aligned}
  \langle T^*Tv, v\rangle &= \langle TT^*v, v\rangle \\
  \langle Tv, Tv\rangle &= \langle T^*v, T^*v\rangle \\
  \|Tv\|^2 &= \|T^*v\|^2 \\
  \|Tv\| &= \|T^*v\|
  \end{aligned}$$

  for all $v \in V$, as desired.

  Now suppose that $\|Tv\| = \|T^*v\|$ for all $v \in V$. Then following the reverse of the procedure for the forward direction, we can easily show that $\langle (T^*T - TT^*)v, v\rangle = 0$ for all $v \in V$. Additionally, by consecutive applications of Theorem 7.2, we have that

  $$\begin{aligned}
  (T^*T - TT^*)^* &= (T^*T)^* - (TT^*)^* \\
  &= T^*(T^*)^* - (T^*)^*T^* \\
  &= T^*T - TT^*
  \end{aligned}$$

  It follows that $T^*T - TT^*$ is self-adjoint. This combined with the previous result implies by Theorem 7.8 that $T^*T - TT^* = 0$. It follows that $T^*T = TT^*$, so T is normal, as desired. ■

- While an operator and its adjoint may have different eigenvectors, a normal operator and its adjoint have the same eigenvectors.

Theorem 7.10. *Suppose $T \in \mathcal{L}(V)$ is normal and $v \in V$ is an eigenvector of T with eigenvalue λ. Then v is also an eigenvector of T^* with eigenvalue $\bar{\lambda}$.*

Proof. By consecutive applications of Theorem 7.2, we have that

$$\begin{aligned}(T-\lambda I)(T-\lambda I)^* &= (T-\lambda I)(T^* - \bar{\lambda} I)\\ &= TT^* - \bar{\lambda}T - \lambda T^* + \lambda\bar{\lambda} I\\ &= T^*T - \lambda T^* - \bar{\lambda}T + \bar{\lambda}\lambda I\\ &= (T^* - \bar{\lambda}I)(T-\lambda I)\\ &= (T-\lambda I)^*(T-\lambda I)\end{aligned}$$

Thus, $T-\lambda I$ is self-adjoint. It follows by Theorem 7.9 that

$$0 = \|(T-\lambda I)v\| = \|(T-\lambda I)^* v\| = \left\|(T^* - \bar{\lambda}I)v\right\|$$

Hence $(T^* - \bar{\lambda}I)v = 0$, so $T^*v = \bar{\lambda}v$, so v is an eigenvector of T^* with eigenvalue $\bar{\lambda}$, as desired. ■

- Normal operators have orthogonal eigenvectors.

Theorem 7.11. *Suppose $T \in \mathcal{L}(V)$ is normal. Then the eigenvectors of T corresponding to distinct eigenvalues are orthogonal.*

Proof. Let α, β be distinct eigenvalues of T, and let u, v be their corresponding eigenvectors. Thus, we have that

$$\begin{aligned}(\alpha-\beta)\langle u,v\rangle &= \langle \alpha u, v\rangle - \langle u, \bar{\beta}v\rangle\\ &= \langle Tu, v\rangle - \langle u, T^*v\rangle && \text{Theorem 7.10}\\ &= 0\end{aligned}$$

Since $\alpha \neq \beta$ by hypothesis, we must have that $\langle u,v\rangle = 0$. Therefore, u, v are orthogonal, as desired. ■

7.B The Spectral Theorem

- Diagonal operators are nice operators.
  - An operator has a diagonal matrix with respect to some basis iff the basis consists of eigenvectors of the operator (see Theorem 5.11).
- The nicest operators are those for which there is an orthonormal basis of V with respect to which the operator has a diagonal matrix.
  - The Spectral Theorem characterizes the operators $T \in \mathcal{L}(V)$ for which there exists an orthonormal basis of V consisting of eigenvectors of T.
  - In particular, it characterizes them as the normal operators when $\mathbb{F} = \mathbb{C}$ and the self-adjoint operators when $\mathbb{F} = \mathbb{R}$.
  - "The Spectral Theorem is probably the most useful tool in the study of operators on inner product spaces" (Axler, 2015, p. 217).
- For the purposes of proving the Spectral Theorem, we will break it into a Complex Spectral Theorem and a Real Spectral Theorem.
- The complex portion is simpler, so we begin with it.

Theorem 7.12 (Complex Spectral Theorem)**.** *Suppoe $\mathbb{F} = \mathbb{C}$ and $T \in \mathcal{L}(V)$. Then the following are equivalent.*

(*a*) *T is normal.*

(*b*) *V has an orthonormal basis consisting of eigenvectors of T.*

(*c*) *T has a diagonal matrix with respect to some orthonormal basis of V.*

Proof. We have by Theorem 5.11 that (b) and (c) are equivalent, so we will focus on proving the equivalence of (a) and (c).

Suppose first that (c) holds. Since $\mathcal{M}(T)$ is diagonal and $\mathcal{M}(T^*)$ is the conjugate transpose of $\mathcal{M}(T)$, $\mathcal{M}(T^*)$ is diagonal. Therefore, since any two diagonal matrices commute, T is normal, so (a) holds.

Now suppose that (a) holds. By Schur's Theorem, there exists an orthonormal basis $e_1, \dots, e_n$ of V with respect to which T has an upper triangular matrix. We will show that this matrix is actually diagonal. To begin, since $\mathcal{M}(T)$ is upper triangular, we know that

$$\|Te_1\|^2 = |a_{1,1}|^2$$

Similarly, since T^* is the conjugate *transpose*, we have that

$$\|T^*e_1\|^2 = |a_{1,1}|^2 + \cdots + |a_{1,n}|^2$$

But since $\|Te_1\| = \|T^*e_1\|$ by Theorem 7.9, the two equations above imply that

$$0 = |a_{1,2}|^2 + \cdots + |a_{1,n}|^2$$

Therefore, we know that all entries in row 1 save the first are zero. We may repeat this procedure for every row to finish the proof. ■

- The next result continues to build on the likeness of normal matrices and real numbers. Specifically, it plays off the fact that if $b, c \in \mathbb{R}$ with $b^2 < 4c$, then $x^2 + bx + c > 0$, i.e., $x^2 + bx + c$ nonzero is an "invertible" real number.

Theorem 7.13. *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 invertible.

Proof. To prove that $T^2 + bT + cI$ is invertible, Theorem 3.19 tells us that it will suffice to show that T is injective. To do this, Theorem 3.4 tells us that we must verify that $\text{null}(T^2 + bT + cI) \subset \{0\}$, i.e., that if $v \in V$ is nonzero, then $(T^2 + bT + cI)v \neq 0$. Let's begin.

Let $v \in V$ be arbitrary. Then we have that

$$\begin{aligned}
\left\langle (T^2 + bT + cI)v, v \right\rangle &= \left\langle T^2 v, v \right\rangle + b \langle Tv, v \rangle + c \langle v, v \rangle \\
&= \langle Tv, Tv \rangle + b \langle Tv, v \rangle + c\|v\|^2 \\
&\geq \|Tv\|^2 - |b|\|Tv\|\|v\| + c\|v\|^2 && \text{Cauchy-Schwarz Inequality} \\
&= \left(\|Tv\| - \frac{|b|\|v\|}{2} \right)^2 + \left(c - \frac{b^2}{4} \right) \|v\|^2 \\
&> 0
\end{aligned}$$

The overall strict inequality implies by the contrapositive of Theorem 6.1b that $(T^2 + bT + cI)v \neq 0$, as desired. ■

- Like Theorem 5.5 told us that operators on *finite-dimensional nonzero complex* vector spaces have eigenvalues, the following tells us that *self-adjoint* operators on *any nonzero* vector space have eigenvalues.

Theorem 7.14. *Suppose $V \neq \{0\}$ and $T \in \mathcal{L}(V)$ is a self-adjoint operator. Then T has an eigenvalue.*

Proof. Let V be a real inner product space, let $n = \dim V$, and let $v \in V$ be arbitrary and nonzero. Since $v, Tv, T^2v, \dots, T^nv$ has length $n+1 > \dim V$, it is linearly dependent. Thus, there exist $a_0, \dots, a_n \in \mathbb{F}$ such that

$$0 = a_0v + a_1Tv + \cdots + a_nT^nv$$

If we let the a's be the coefficients of a degree n polynomial, then we have by Theorem 4.9 that

$$a_0 + a_1x + \cdots + a_nx^n = c(x^2 + b_1x + c_1) \cdots (x^2 + b_Mx + c_M)(x - \lambda_1) \cdots (x - \lambda_m)$$

where $c \in \mathbb{R}$ is nonzero, each $b_j, c_j, \lambda_j \in \mathbb{R}$, each $b_j^2 < 4c_j$, $m + M \geq 1$, and the equation holds for all $x \in \mathbb{R}$. It follows that

$$\begin{aligned}
0 &= a_0v + a_1Tv + \cdots + a_nT^nv \\
&= (a_0I + a_1T + \cdots + a_nT^n)v \\
&= c(T^2 + b_1T + c_1I) \cdots (T_2 + b_MT + c_MI)(T - \lambda_1I) \cdots (T - \lambda_mI)v
\end{aligned}$$

Since T is self-adjoint and $b_j, c_j \in \mathbb{R}$ satisfy $b_j^2 < 4c_j$ for each j, we have by consecutive applications of Theorem 7.13 that each $T^2 + b_jT + c_jI$ is invertible. Thus, if we multiply both sides of the above equation by $1/c$ (recall that $c \neq 0$) and $(T^2 + b_jT + c_jI)^{-1}$ for each j, we obtain

$$0 = (T - \lambda I) \cdots (T - \lambda_mI)v$$

Therefore, by an argument symmetric to that used in the last paragraph of the proof of Theorem 5.5, we have that T has an eigenvalue, as desired. ∎

- Invariant subspaces and self-adjoint operators.

Theorem 7.15. *Suppose $T \in \mathcal{L}(V)$ is self-adjoint and U is a subspace of V that is invariant under T. Then*

(a) *$U^\perp$ is invariant under T.*

Proof. Let $v \in U^\perp$ be arbitrary, and let u be any element of U. Then

$$\langle Tv, u \rangle = \langle v, Tu \rangle = 0$$

where the first equality holds because T is self-adjoint and the second equality holds because U is invariant under T (so $Tu \in U$, and we know that the inner product of an element of $U^\perp$ with an element of U is 0). Thus, since $\langle Tv, u \rangle = 0$ for all $u \in U$, $Tv \in U^\perp$, as desired. ∎

(b) *$T|_U \in \mathcal{L}(U)$ is self-adjoint.*

Proof. If $u, v \in U$, then

$$\langle (T|_U)u, v \rangle = \langle Tu, v \rangle = \langle u, Tv \rangle = \langle u, (T|_U)v \rangle$$

as desired. ∎

(c) *$T|_{U^\perp} \in \mathcal{L}(U^\perp)$ is self-adjoint.*

Proof. The proof is symmetric to that of Theorem 7.15b. ∎

- We can now prove the real portion of the spectral theorem.

Theorem 7.16 (Real Spectral Theorem)**.** *Suppose $\mathbb{F} = \mathbb{R}$ and $T \in \mathcal{L}(V)$. Then the following are equivalent.*

(*a*) *T is self-adjoint.*

(*b*) *V has an orthonormal basis consisting of eigenvectors of T.*

(*c*) *T has a diagonal matrix with respect to some orthonormal basis of V.*

Proof. We will prove that (a) implies (b), (b) implies (c), and (c) implies (a). Let's begin.

First, suppose that T is self-adjoint. We induct on $\dim V$. For the base case $\dim V = 1$, we must have $Tv = \lambda v$ for any $v \in V$. Thus, take $e = v/\|v\|$ as an orthonormal basis of V consisting of eigenvectors of T. Now suppose inductively that (a) implies (b) for all real inner product spaces of dimension less than $\dim V > 1$. Suppose $T \in \mathcal{L}(V)$ is self-adjoint. By Theorem 7.14, we may let v be an eigenvector of T. It follows that $u = v/\|v\|$ is a normal eigenvector of T. Let $U = \text{span}(u)$. Then U is a subspace of V that is invariant under T, so we have by Theorem 7.15c that $T|_{U^\perp} \in \mathcal{L}(U^\perp)$ is self-adjoint. But since $\dim U^\perp = \dim V - \dim U = \dim V - 1$, we have by the inductive hypothesis that there is an orthonormal basis of $U^\perp$ consisting of eigenvectors of $T|_{U^\perp}$. Adjoining u to this list gives an orthonormal basis of V consisting of eigenvectors of T, as desired.

Second, suppose that V has an orthonormal basis $e_1, \dots, e_n$ consisting of eigenvectors of T. Then since

$$Te_j = 0e_1 + \cdots + 0e_{j-1} + \lambda_j e_j + 0e_{j+1} + \cdots + 0e_n$$

for all j, we have by the definition that $\mathcal{M}(T, (e_1, \dots, e_n))$ is diagonal, as desired.

Third, suppose that T has a diagonal matrix $\mathcal{M}(T)$ with respect to some orthonormal basis of V. In a real inner product space, $\overline{\mathcal{M}(T)} = \mathcal{M}(T)$. Additionally, any diagonal matrix is equal to its transpose. Thus, $T = T^*$, so T is self-adjoint, as desired. ∎

7.C Positive Operators and Isometries

10/11: • **Positive** ($T \in \mathcal{L}(V)$): A self-adjoint operator $T \in \mathcal{L}(V)$ such that

$$\langle Tv, v \rangle \geq 0$$

for all $v \in V$. *Also known as* **positive semidefinite** (operator).

- Note that if V is complex, Theorem 7.7 implies based on the condition that $\langle Tv, v \rangle \geq 0$ for all $v \in V$ that T is self-adjoint. Therefore, in this case, we need not explicitly postulate that T is self-adjoint.

• **Square root** (of $T \in \mathcal{L}(V)$): An operator R such that $R^2 = T$.

• The following characterization of positive operators is directly analogous to the characterization of nonnegative complex numbers.

Theorem 7.17. *Let $T \in \mathcal{L}(V)$. Then the following are equivalent.*

(*a*) *T is positive.*

(*b*) *T is self-adjoint and all the eigenvalues of T are nonnegative.*

(*c*) *T has a positive square root.*

(*d*) *T has a self-adjoint square root.*

(*e*) *There exists an operator $R \in \mathcal{L}(V)$ such that $T = R^*R$.*

Proof. We will prove that (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (e) $\Rightarrow$ (a). Let's begin.

First, suppose that T is positive. Then by definition, T is self-adjoint. Additionally, let $\lambda \in \mathbb{F}$ be an eigenvalue of T. It follows by the definition of positive operators and by the positivity of the inner product that

$$\begin{aligned} 0 &\leq \langle Tv, v \rangle = \langle \lambda v, v \rangle = \lambda \langle v, v \rangle \\ 0 &\leq \lambda \end{aligned}$$

as desired.

Second, suppose that T is self-adjoint and all the eigenvalues of T are nonnegative. Since T is self-adjoint, the Real and Complex Spectral Theorems imply that there exists an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of T. Let $\lambda_1, \dots, \lambda_n$ be the corresponding eigenvalues; it follows by hypothesis that $\lambda_j \geq 0$ for all j. We now define $R \in \mathcal{L}(V)$ by

$$Re_j = \sqrt{\lambda_j} e_j$$

for all j. To prove that R is positive, let $v \in V$ be arbitrary. Suppose $v = a_1 e_1 + \cdots + a_n e_n$ where $a_1, \dots, a_n \in \mathbb{F}$. Then

$$\begin{aligned}
\langle Rv, v \rangle &= \langle R(a_1 e_1 + \cdots + a_n e_n), a_1 e_1 + \cdots + a_n e_n \rangle \\
&= \left\langle \sqrt{\lambda_1} a_1 e_1 + \cdots + \sqrt{\lambda_n} a_n e_n, a_1 e_1 + \cdots + a_n e_n \right\rangle \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \left\langle \sqrt{\lambda_i} a_i e_i, a_j e_j \right\rangle \\
&= \sum_{i=1}^{n} \left\langle \sqrt{\lambda_i} a_i e_i, a_i e_i \right\rangle \\
&= \sum_{i=1}^{n} \sqrt{\lambda_i} \\
&\geq 0
\end{aligned}$$

as desired. Furthermore, $R^2 e_j = \lambda_j e_j = Te_j$ for each j, so by Theorem 3.1, $R^2 = T$, as desired.

Third, suppose that T has a positive square root R. Then by the definition of a positive operator, R is self-adjoint as well, as desired.

Fourth, suppose that T has a self-adjoint square root R. Since R is self-adjoint, $R = R^*$. Therefore,

$$T = R^2 = R^* R$$

as desired.

Fifth, suppose that there exists an operator $R \in \mathcal{L}(V)$ such that $T = R^* R$. To prove that T is positive, it will suffice to show that it is self-adjoint and that $\langle Tv, v \rangle \geq 0$ for all $v \in V$. First off, T is self-adjoint since

$$T^* = (R^* R)^* = R^* (R^*)^* = R^* R = T$$

Second, we have that

$$\langle Tv, v \rangle = \langle R^* Rv, v \rangle = \langle Rv, Rv \rangle \geq 0$$

for all $v \in V$. Therefore, T is positive, as desired. ■

- Since each nonnegative number has a unique nonnegative square root, the next result makes sense by analogy.

Theorem 7.18. *Every positive operator on V has a unique positive square root.*

Proof. Let T be a positive operator on V, let $v \in V$ be an eigenvector of T, let $\lambda \in \mathbb{F}$ be the corresponding eigenvalue, and let R be a positive square root of T (Theorem 7.17 guarantees that at least one such operator exists). Since T is positive, Theorem 7.17 implies that $\lambda \geq 0$. Thus, to prove that R is unique, we will prove that $Rv = \sqrt{\lambda} v$. This will imply that the behavior of R on the eigenvectors of T is uniquely determined. It will follows since there is a basis of V consisting of the eigenvectors of T (by the Real and Complex Spectral Theorems), the behavior of R on V (and hence R) is uniquely determined. Let's begin.

Since R is positive (hence self-adjoint), the Real and Complex Spectral Theorems assert that there exists an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of R. Additionally, because R is positive, the corresponding eigenvalues $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}$ are nonnegative.

Now let

$$v = a_1 e_1 + \cdots + a_n e_n$$

for $a_1, \dots, a_n \in \mathbb{F}$. Then

$$a_1 \lambda e_1 + \cdots + a_n \lambda e_n = Tv = R^2 v = a_1 \lambda_1 e_1 + \cdots + a_n \lambda_n e_n$$

so since $e_1, \dots, e_n$ is linearly independent, $a_j(\lambda - \lambda_j) = 0$ for all j. It follows that

$$v = \sum_{\{j:\lambda_j = \lambda\}} a_j e_j$$

so that

$$\begin{aligned}
Rv &= \sum_{j=1}^{n} a_j \sqrt{\lambda_j} e_j \\
&= \sum_{\{j:\lambda_j = \lambda\}} a_j \sqrt{\lambda_j} e_j \\
&= \sum_{\{j:\lambda_j = \lambda\}} a_j \sqrt{\lambda} e_j \\
&= \sqrt{\lambda} v
\end{aligned}$$

as desired. ■

- **Isometry**: An operator $S \in \mathcal{L}(V)$ such that

  $$\|Sv\| = \|v\|$$

  for all $v \in V$.
  - In other words, an isometry is an operator that preserves norms.
- **Orthogonal** (operator): An isometry on a real inner product space.
- **Unitary** (operator): An isometry on a complex inner product space.
- Characterizing isometries.

  Theorem 7.19. *Suppose $S \in \mathcal{L}(V)$. Then the following are equivalent.*

  (a) *S is an isometry.*
  (b) *$\langle Su, Sv \rangle = \langle u, v \rangle$ for all $u, v \in V$.*
  (c) *$Se_1, \dots, Se_n$ is orthonormal for every orthonormal list of vectors $e_1, \dots, e_n$ in V.*
  (d) *There exists an orthonormal basis $e_1, \dots, e_n$ of V such that $Se_1, \dots, Se_n$ is orthonormal.*
  (e) *$S^*S = I$.*
  (f) *$SS^* = I$.*
  (g) *S^* is an isometry.*
  (h) *S is invertible and $S^{-1} = S^*$.*

Proof. We will prove that (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (e) $\Rightarrow$ (f) $\Rightarrow$ (g) $\Rightarrow$ (h) $\Rightarrow$ (a). Let's begin.

First, suppose that S is an isometry. Let $u, v \in V$ be arbitrary. We divide into two cases (V is a real inner product space and V is a complex inner product space). If V is a real inner product space, then

$$\begin{aligned}
\langle Su, Sv \rangle &= \frac{\|Su + Sv\|^2 - \|Su - Sv\|^2}{4} && \text{Exercise 6.A.19} \\
&= \frac{\|S(u+v)\|^2 - \|S(u-v)\|^2}{4} \\
&= \frac{\|u+v\|^2 - \|u-v\|^2}{4} \\
&= \langle u, v \rangle && \text{Exercise 6.A.19}
\end{aligned}$$

as desired. On the other hand, if V is a complex vector space, then the proof is symmetric to the above except with the use of Exercise 6.A.20 instead of Exercise 6.A.19.

Second, suppose that $\langle Su, Sv \rangle = \langle u, v \rangle$ for all $u, v \in V$. Let $e_1, \dots, e_n$ be an orthonormal list of vectors in V. Then by hypothesis,

$$\langle Se_i, Se_j \rangle = \langle e_i, e_j \rangle = \delta_{ij}$$

for all $1 \leq i, j \leq n$, proving that $Se_1, \dots, Se_n$ is orthonormal, as desired.

Third, suppose that $Se_1, \dots, Se_n$ is orthonormal for every orthonormal list of vectors $e_1, \dots, e_n$ in V. By Theorem 6.14, there exists an orthonormal basis $e_1, \dots, e_n$ of V. It follows by hypothesis that $Se_1, \dots, Se_n$ is orthonormal, as desired.

Fourth, suppose that there exists an orthonormal basis $e_1, \dots, e_n$ of V such that $Se_1, \dots, Se_n$ is orthonormal. Then

$$\langle S^* Se_j, e_k \rangle = \langle Se_j, Se_k \rangle = \delta_{jk} = \langle e_j, e_k \rangle$$

for all $1 \leq j, k \leq n$. It follows that if $u, v \in V$, then

$$\begin{aligned}
\langle S^* Su, v \rangle &= \langle S^* S(a_1 e_1 + \cdots + a_n e_n), b_1 e_1 + \cdots + b_n e_n \rangle \\
&= \langle S^* S a_1 e_1, b_1 e_1 \rangle + \cdots + \langle S^* S a_n e_n, b_n e_n \rangle \\
&= \langle a_1 e_1, b_1 e_1 \rangle + \cdots + \langle a_n e_n, b_n e_n \rangle \\
&= \langle a_1 e_1 + \cdots + a_n e_n, b_1 e_1 + \cdots + b_n e_n \rangle \\
&= \langle u, v \rangle
\end{aligned}$$

Therefore, $S^* S = I$, as desired.

Fifth, suppose that $S^* S = I$. Then by Exercise 3.D.10, $SS^* = I$, as desired.

Sixth, suppose that $SS^* = I$. To prove that S^* is an isometry, it will suffice to show that $\|S^* v\| = \|v\|$ for all $v \in V$. Let $v \in V$ be arbitrary. Then

$$\|S^* v\|^2 = \langle S^* v, S^* v \rangle = \langle SS^* v, v \rangle = \langle v, v \rangle = \|v\|^2$$

Taking square roots yields the desired equality.

Seventh, suppose that S^* is an isometry. It follows by our previous chain of proofs that $(S^*)^* S^* = SS^* = I$ and $S^* (S^*)^* = S^* S = I$. Therefore, S is invertible with inverse $S^{-1} = S^*$, as desired.

Eighth, suppose that S is invertible and $S^{-1} = S^*$. Then if $v \in V$, we have that

$$\|Sv\|^2 = \langle Sv, Sv \rangle = \langle S^* Sv, v \rangle = \langle S^{-1} Sv, v \rangle = \langle v, v \rangle = \|v\|^2$$

Taking square roots yields the desired equality. ∎

- It follows from (e) and (f) that every isometry is normal.

- Thus, characterizations of normal operators (e.g., the Complex Spectral Theorem) can be used to give descriptions of isometries.

Theorem 7.20. *Suppose V is a complex inner product space and $S \in \mathcal{L}(V)$. Then the following are equivalent.*

(a) *S is an isometry.*

(b) *There is an orthonormal basis of V consisting of eigenvectors of S whose corresponding eigenvalues all have absolute value 1.*

Proof. Suppose first that S is an isometry. Then by the Complex Spectral Theorem, there is an orthonormal basis $e_1, \dots, e_n$ of V consisting of the eigenvectors of S. Let $\lambda_1, \dots, \lambda_n$ be the corresponding eigenvalues. Then for each $j = 1, \dots, n$, we have that

$$|\lambda_j| = \|\lambda_j e_j\| = \|Se_j\| = \|e_j\| = 1$$

as desired.

Now suppose that there is an orthonormal basis $e_1, \dots, e_n$ of V consisting of eigenvectors of S whose corresponding eigenvalues $\lambda_1, \dots, \lambda_n$ all have absolute value 1. Let $v \in V$ be arbitrary. Then by Theorem 6.12, we have that

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n \qquad \|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2$$

It follows that

$$\begin{aligned} Sv &= \langle v, e_1 \rangle Se_1 + \cdots + \langle v, e_n \rangle Se_n \\ &= \langle v, e_1 \rangle \lambda_1 e_1 + \cdots + \langle v, e_n \rangle \lambda_n e_n \end{aligned}$$

Thus, we have that

$$\begin{aligned} \|Sv\|^2 &= \langle \langle v, e_1 \rangle \lambda_1 e_1 + \cdots + \langle v, e_n \rangle \lambda_n e_n, \langle v, e_1 \rangle \lambda_1 e_1 + \cdots + \langle v, e_n \rangle \lambda_n e_n \rangle \\ &= \langle \langle v, e_1 \rangle \lambda_1 e_1, \langle v, e_1 \rangle \lambda_1 e_1 \rangle + \cdots + \langle \langle v, e_n \rangle \lambda_n e_n, \langle v, e_n \rangle \lambda_n e_n \rangle \\ &= \langle v, e_1 \rangle \lambda_1 \cdot \overline{\langle v, e_1 \rangle \lambda_1} \cdot \langle e_1, e_1 \rangle + \cdots + \langle v, e_n \rangle \lambda_n \cdot \overline{\langle v, e_n \rangle \lambda_n} \cdot \langle e_n, e_n \rangle \\ &= \langle v, e_1 \rangle \lambda_1 \cdot \overline{\langle v, e_1 \rangle} \bar{\lambda}_1 \cdot 1 + \cdots + \langle v, e_n \rangle \lambda_n \cdot \overline{\langle v, e_n \rangle} \bar{\lambda}_n \cdot 1 \\ &= |\langle v, e_1 \rangle|^2 |\bar{\lambda}_1|^2 + \cdots + |\langle v, e_n \rangle|^2 |\bar{\lambda}_n|^2 \\ &= |\langle v, e_1 \rangle|^2 \cdot 1 + \cdots + |\langle v, e_n \rangle|^2 \cdot 1 \\ &= \|v\|^2 \end{aligned}$$

Taking square roots yields the desired equality. ∎

7.D Polar Decomposition and Singular Value Decomposition

10/16: - **Square root** (of a positive $T \in \mathcal{L}(V)$): The unique positive operator $R \in \mathcal{L}(V)$ such that $R^2 = T$. *Denoted by* $\boldsymbol{\sqrt{T}}$.
  - The existence of such an operator is justified by Theorem 7.18.

10/18: - Continuing with our analogy between $\mathbb{C}$ and $\mathcal{L}(V)$, we now prove an analogous theorem to the decomposition of any complex number z into the form $z = (z/|z|)|z| = (z/|z|)\sqrt{\bar{z}z}$, where $z/|z|$ (as an element of the unit circle) is analogous to an isometry, and $\bar{z}$ is analogous to the adjoint.

Theorem 7.21 (Polar Decomposition)**.** *Suppose $T \in \mathcal{L}(V)$. Then there exists an isometry $S \in \mathcal{L}(V)$ such that*

$$T = S\sqrt{T^*T}$$

Lemma. *If $v \in V$, then*

$$\|Tv\| = \left\|\sqrt{T^*T}v\right\|$$

Proof. Let $v \in V$ be arbitrary. Then

$$\begin{aligned}\|Tv\|^2 &= \langle Tv, Tv\rangle \\ &= \langle T^*Tv, v\rangle \\ &= \left\langle \sqrt{T^*T}\sqrt{T^*T}v, v\right\rangle \\ &= \left\langle \sqrt{T^*T}v, \sqrt{T^*T}v\right\rangle \\ &= \left\|\sqrt{T^*T}v\right\|^2\end{aligned}$$

where the third equality holds because T^*T is positive by Theorem 7.17 and thus has a positive square root, and the fourth equality holds because $\sqrt{T^*T}$ is positive and thus is self-adjoint by definition. Taking square roots of the above gives the desired inequality. ■

Proof of Theorem 7.21. For this proof, we will first define a map $S_1 : \operatorname{range}\sqrt{T^*T} \to \operatorname{range} T$. We will then prove that it is a well-defined function and that it is a linear map. S_1 thus has the desired property; all that remains is to extend it to an isometry. To do so, we define $S_2 : (\operatorname{range}\sqrt{T^*T})^\perp \to (\operatorname{range} T)^\perp$ so that S, defined as S_1 on the appropriate domain and S_2 on its complement, is an isometry. Let's begin.

Let $S_1 : \operatorname{range}\sqrt{T^*Tv} \to \operatorname{range} T$ be defined by

$$S_1(\sqrt{T^*T}v) = Tv$$

for all $\sqrt{T^*T}v \in \operatorname{range}\sqrt{T^*T}$.

To prove that S_1 is a function, it will suffice to show that if $\sqrt{T^*T}v_1 = \sqrt{T^*T}v_2$, then $Tv_1 = Tv_2$. But if $\sqrt{T^*T}v_1 = \sqrt{T^*T}v_2$, then

$$\begin{aligned}\|Tv_1 - Tv_2\| &= \|T(v_1 - v_2)\| \\ &= \left\|\sqrt{T^*T}(v_1 - v_2)\right\| && \text{Lemma} \\ &= \left\|\sqrt{T^*T}v_1 - \sqrt{T^*T}v_2\right\| \\ &= 0\end{aligned}$$

Thus, by Theorem 6.2a, $Tv_1 - Tv_2 = 0$, so $Tv_1 = Tv_2$, as desired.

To prove that S_1 is a linear map, it will suffice to show that $S_1(\alpha\sqrt{T^*T}v) = \alpha S_1(\sqrt{T^*T}v)$ where $\alpha \in \mathbb{F}$ and $S_1(\sqrt{T^*T}v_1 + \sqrt{T^*T}v_2) = S_1(\sqrt{T^*T}v_1) + S_1(\sqrt{T^*T}v_2)$. But since $\sqrt{T^*T}$ and T are both linear maps themselves, we have that

$$\begin{aligned}S_1(\alpha\sqrt{T^*T}v) &= S_1(\sqrt{T^*T}(\alpha v)) \\ &= T(\alpha v) \\ &= \alpha Tv \\ &= \alpha S_1(\sqrt{T^*T}v)\end{aligned} \qquad \begin{aligned}S_1(\sqrt{T^*T}v_1 + \sqrt{T^*T}v_2) &= S_1(\sqrt{T^*T}(v_1 + v_2)) \\ &= T(v_1 + v_2) \\ &= Tv_1 + Tv_2 \\ &= S_1(\sqrt{T^*T}v_1) + S_1(\sqrt{T^*T}v_2)\end{aligned}$$

as desired.

To prove that S_1 is an isometry, it will suffice to show that $\|S_1(\sqrt{T^*T}v)\| = \|\sqrt{T^*T}v\|$ for all $\sqrt{T^*T}v \in \operatorname{range}\sqrt{T^*T}$. But if $\sqrt{T^*T}v \in \operatorname{range}\sqrt{T^*T}$, then

$$\begin{aligned}\left\|S_1(\sqrt{T^*T}v)\right\| &= \|Tv\| \\ &= \left\|\sqrt{T^*T}v\right\|\end{aligned}$$

where the first equality holds by the definition of S_1, and the second holds by the Lemma.

We now build up to our definition of S_2. For starters, notice that it follows from the Lemma that S_1 is injective much the same way it followed that S_1 was a function. Consequently, Theorem 3.4 asserts that $\operatorname{null} S_1 = \{0\}$. Thus, since $S_1 \in \mathcal{L}(\operatorname{range}\sqrt{T^*T}, \operatorname{range} T)$, we have by the Fundamental Theorem of Linear Maps that

$$\begin{aligned}\dim\operatorname{range}\sqrt{T^*T} &= \dim\operatorname{null} S_1 + \dim\operatorname{range} S_1\\ &= 0 + \dim\operatorname{range} T\\ &= \dim\operatorname{range} T\end{aligned}$$

Thus, since $\operatorname{range}\sqrt{T^*T} \subset V$ and $\operatorname{range} T \subset V$, we have that

$$\begin{aligned}\dim(\operatorname{range} T)^\perp &= \dim V - \dim\operatorname{range} T && \text{Theorem 6.21}\\ &= \dim V - \dim\operatorname{range}\sqrt{T^*T}\\ &= \dim(\operatorname{range}\sqrt{T^*T})^\perp && \text{Theorem 6.21}\end{aligned}$$

It follows that we can choose orthonormal bases $e_1, \dots, e_m$ and $f_1, \dots, f_m$ of $(\operatorname{range}\sqrt{T^*T})^\perp$ and $(\operatorname{range} T)^\perp$ of equal length. Let $S_2 : (\operatorname{range}\sqrt{T^*T})^\perp \to (\operatorname{range} T)^\perp$ be the unique linear transformation such that $Te_j = f_j$ for each $j = 1, \dots, n$ implied to exist by Theorem 3.1. Note that S_2 is also an isometry since if $x \in (\operatorname{range}\sqrt{T^*T})^\perp$, then

$$\begin{aligned}\|S_2x\|^2 &= \|S_1(a_1e_1 + \cdots + a_me_m)\|^2\\ &= \|a_1f_1 + \cdots + a_mf_m\|^2\\ &= |a_1|^2 + \cdots + |a_m|^2 && \text{Theorem 6.9}\\ &= \|a_1e_1 + \cdots + a_mf_m\|^2 && \text{Theorem 6.9}\\ &= \|x\|^2\end{aligned}$$

where taking square roots yields the desired equality.

We are now ready to define $S \in \mathcal{L}(V)$. Let $v \in V$ be arbitrary. It follows by Theorem 6.20 that we can uniquely decompose v into a sum $v = u + w$ where $u \in \operatorname{range}\sqrt{T^*T}$ and $v \in (\operatorname{range}\sqrt{T^*T})^\perp$. Thus, we define

$$Sv = S_1u + S_2w$$

We could (but will not) explicitly show based on the previously proven properties that S is well-defined and linear. We will, however, show that S is an isometry: for any $v \in V$,

$$\begin{aligned}\|Sv\|^2 &= \|S_1u + S_2v\|^2\\ &= \|S_1u\|^2 + \|S_2w\| && \text{Pythagorean Theorem}\\ &= \|u\|^2 + \|w\|^2\\ &= \|v\|^2 && \text{Pythagorean Theorem}\end{aligned}$$

Lastly, we have by its definition that for any $v \in V$,

$$(S\sqrt{T^*T})v = S(\sqrt{T^*T}v) = S_1(\sqrt{T^*T}v) = Tv$$

so $T = S\sqrt{T^*T}$, as desired. ■

10/21:

- The main conclusion from the Polar Decomposition is that *any* linear operator, no matter how ill-defined, can be decomposed into the product of an isometry and a positive operator, two very well characterized classes of operators.

  - In particular, if $\mathbb{F} = \mathbb{C}$, then T is the product of two operators, both of which are orthonormally diagonalizable (though not necessarily with respect to the same orthonormal bases).

- **Singular values** (of $T \in \mathcal{L}(V)$): The eigenvalues of $\sqrt{T^*T}$, with each value λ repeated $\dim E(\lambda, \sqrt{T^*T})$ times.
- The singular values of T are all nonnegative (because $\sqrt{T^*T}$ is a positive operator [see Theorem 7.17]).
- Each $T \in \mathcal{L}(V)$ has $\dim V$ singular values (because $\sqrt{T^*T}$ is positive, hence self-adjoint, hence $\sqrt{T^*T}$ has a diagonal matrix [see the Real Spectral Theorem], hence $\sqrt{T^*T}$ has $\dim V$ distinct eigenvalues).
- We now show that every operator on V can be described in terms of its singular values and two orthonormal bases on V.

Theorem 7.22 (Singular Value Decomposition)**.** *Suppose $T \in \mathcal{L}(V)$ has singular values $s_1, \dots, s_n$. Then there exist orthonormal bases $e_1, \dots, e_n$ and $f_1, \dots, f_n$ of V such that*

$$Tv = s_1 \langle v, e_1 \rangle f_1 + \cdots + s_n \langle v, e_n \rangle f_n$$

for every $v \in V$.

Proof. Applying the Real Spectral Theorem to the self-adjoint operator $\sqrt{T^*T}$ reveals that V has an orthonormal basis $e_1, \dots, e_n$ of eigenvectors of $\sqrt{T^*T}$. Therefore, if we let $v \in V$ be arbitrary, then we have that

$$\begin{aligned}
Tv &= (S\sqrt{T^*T})v && \text{Polar Decomposition} \\
&= S(\sqrt{T^*T}(\langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n)) && \text{Theorem 6.12} \\
&= S(\langle v, e_1 \rangle \sqrt{T^*T} e_1 + \cdots + \langle v, e_n \rangle \sqrt{T^*T} e_n) \\
&= S(\langle v, e_1 \rangle s_1 e_1 + \cdots + \langle v, e_n \rangle s_n e_n) \\
&= s_1 \langle v, e_1 \rangle S e_1 + \cdots + s_n \langle v, e_n \rangle S e_n \\
&= s_1 \langle v, e_1 \rangle f_1 + \cdots + s_n \langle v, e_n \rangle f_n && \text{Theorem 7.19}
\end{aligned}$$

where $s_1, \dots, s_n$ are the singular values of T (the eigenvalues of $\sqrt{T^*T}$) and $f_1, \dots, f_n$ is another orthonormal basis of V. ■

- If $e_1, \dots, e_n$ and $f_1, \dots, f_n$ are orthonormal bases of V that satisfy the Singular Value Decomposition for some operator T, then $Te_j = s_j f_j$ for each $j = 1, \dots, n$.
  - In other words, every operator on V has a diagonal matrix with respect to some orthonormal *bases* (plural) of V.
- The Singular Value Decomposition has many applications, especially in the realm of computational linear algebra, where working with T^*T is much easier than working with $\sqrt{T^*T}$. A powerful tool in this pursuit is the following.

Theorem 7.23. *Suppose $T \in \mathcal{L}(V)$. Then the singular values of T are the nonnegative square roots of the eigenvalues of T^*T, with each eigenvalue λ repeated $\dim E(\lambda, T^*T)$ times.*

Proof. Since T^*T is positive and self-adjoint, we have by the hyperref[trm:RealSpectral]Real Spectral Theorem that there exists an orthonormal basis $e_1, \dots, e_n$ of V and nonnegative numbers $\lambda_1, \dots, \lambda_n$ such that $T^*Te_j = \lambda_j e_j$ for each $j = 1, \dots, n$. It follows since $\sqrt{T^*T}$ is also a positive, self-adjoint operator that its eigenvalues (which exist) must be of the form $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}$ to satisfy $(\sqrt{T^*T})^2 = T^*T$ and to be nonnegative. ■

Chapter 8

Operators on Complex Vector Spaces

8.A Generalized Eigenvectors and Nilpotent Operators

10/22:

- In this chapter, we will assume that V is a finite-dimensional *nonzero* vector space over $\mathbb{F}$ (just to avoid dealing with some trivialities).
- Null spaces and powers of an operator.

Theorem 8.1. *Suppose $T \in \mathcal{L}(V)$. Then*

$$\{0\} = \operatorname{null} T^0 \subset \operatorname{null} T^1 \subset \cdots$$

Proof. We induct on the exponent k of T. For the base case $k = 0$, suppose $v \in \operatorname{null} T^0$. Then $v \in \operatorname{null} I$ since $T^0 = I$ by definition. It follows that

$$0 = Iv = v$$

so $\{0\} = \operatorname{null} T^0$, as desired. Now suppose inductively that we have proven the claim for k; we now wish to show that $\operatorname{null} T^k \subset \operatorname{null} T^{k+1}$. Suppose $v \in \operatorname{null} T^k$. Then $T^k v = 0$. It follows that

$$T^{k+1} v = T(T^k v) = T(0) = 0$$

so $v \in T^{k+1}$, as desired. ∎

Theorem 8.2. *Let $T \in \mathcal{L}(V)$, and suppose m is a nonnegative integer such that* $\operatorname{null} T^m = \operatorname{null} T^{m+1}$. *Then*

$$\operatorname{null} T^m = \operatorname{null} T^{m+1} = \operatorname{null} T^{m+2} = \cdots$$

Proof. We induct on k, defined as follows. For the base case $k = 0$, we have that

$$\operatorname{null} T^{m+0} = \operatorname{null} T^m = \operatorname{null} T^{m+1} = \operatorname{null} T^{m+0+1}$$

by hypothesis, as desired. Now suppose inductively that we have proven that $\operatorname{null} T^{m+k-1} = \operatorname{null} T^{m+k}$; we now wish to show that $\operatorname{null} T^{m+k} = \operatorname{null} T^{m+k+1}$. By Theorem 8.1, we have that $\operatorname{null} T^{m+k} \subset \operatorname{null} T^{m+k+1}$. On the other hand, suppose that $v \in \operatorname{null} T^{m+k+1}$. Then

$$0 = T^{m+k+1} v = T^{m+1}(T^k v)$$

But this implies that $T^k v \in \operatorname{null} T^{m+1} = \operatorname{null} T^m$ by hypothesis. Therefore,

$$0 = T^m(T^k v) = T^{m+k} v$$

so $v \in \operatorname{null} T^{m+k}$, as desired. ∎

- Theorem 8.2 raises the question how to characterize/define/find nonnegative integers m such that the null space stops growing. We tackle begin to tackle this question with the following.

  Theorem 8.3. *Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then*

  $$\operatorname{null} T^n = \operatorname{null} T^{n+1} = \cdots$$

  Proof. To prove the claim, Theorem 8.2 tells us that we need only verify that $\operatorname{null} T^n = \operatorname{null} T^{n+1}$. Suppose for the sake of contradiction that $\operatorname{null} T^n \neq \operatorname{null} T^{n+1}$. Then by Theorem 8.2, we cannot have $\operatorname{null} T^k = \operatorname{null} T^{k+1}$ for any $0 \leq k \leq n$. However, by Theorem 8.1, we must still have that $\operatorname{null} T^k \subset \operatorname{null} T^{k+1}$ for each $k = 1, \dots, n$. Combining the last two results, we must have the following.

  $$\{0\} = \operatorname{null} T^0 \subsetneq \operatorname{null} T^1 \subsetneq \cdots \subsetneq \operatorname{null} T^n \subsetneq \operatorname{null} T^{n+1}$$

  At each of these strict inclusions, the dimension from the previous to the next null space must increase by at least one. Thus, $\dim \operatorname{null} T^{n+1} \geq n+1$. But since $\operatorname{null} T^{n+1} \subset V$, Theorem 2.11 asserts that $\dim \operatorname{null} T^{n+1} \leq n$, so we have that

  $$n + 1 \leq \dim \operatorname{null} T^{n+1} \leq n$$

  a contradiction. ■

- While it is not true that $V = \operatorname{null} T \oplus \operatorname{range} T$ for each $T \in \mathcal{L}(V)$, we can prove the following related theorem.

  Theorem 8.4. *Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then*

  $$V = \operatorname{null} T^n \oplus \operatorname{range} T^n$$

  Proof. To prove that $V = \operatorname{null} T^n \oplus \operatorname{range} T^n$, it will suffice to show that $(\operatorname{null} T^n) \cap (\operatorname{range} T^n) = \{0\}$ and that $\dim(\operatorname{null} T^n \oplus \operatorname{range} T^n) = \dim V$ (see Exercise 2.C.1). Let's begin.

  Suppose $v \in (\operatorname{null} T^n) \cap (\operatorname{range} T^n)$. Then $T^n v = 0$ and there exists $u \in V$ such that $v = T^n u$. Combining these results reveals that

  $$T^{2n} u = T^n v = 0$$

  so $u \in \operatorname{null} T^{2n} = \operatorname{null} T^n$ by Theorem 8.3. Therefore, $v = T^n u = 0$, as desired.

  As to the other equality, we have that

  $$\begin{aligned} \dim(\operatorname{null} T^n \oplus \operatorname{range} T^n) &= \dim \operatorname{null} T^n + \dim \operatorname{range} T^n && \text{Theorem 3.22} \\ &= \dim V && \text{Fundamental Theorem of Linear Maps} \end{aligned}$$

  as desired. ■

- Although many operators can be described by their eigenvectors, not all can. Thus, we introduce the following more general descriptor.

- **Generalized eigenvector** (of $T \in \mathcal{L}(V)$): A nonzero vector $v \in V$ such that

  $$(T - \lambda I)^j v = 0$$

  for some positive integer j, where λ is an eigenvalue of T.

  – Although this definition lets j be arbitrary, we will soon prove that if $j = \dim V$, every generalized eigenvector satisfies the above equation.

  – Note that we do not define generalized eigenvalues because generalized eigenvectors still pertain to the original set of eigenvalues.

- Every eigenvector of T is a generalized eigenvector of T (take $j = 1$ in the definition).
- **Generalized eigenspace** (of $T \in \mathcal{L}(V)$ and λ): The set of all generalized eigenvectors of T corresponding to λ, and the 0 vector. *Denoted by* $\boldsymbol{G(\lambda, T)}$.
- Since every eigenvector of T is a generalized eigenvector of T, we have that $E(\lambda, T) \subset G(\lambda, T)$.
- We now characterize generalized eigenspaces.

Theorem 8.5. *Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. Then $G(\lambda, T) = \text{null}(T - \lambda I)^{\dim V}$.*

Proof. Suppose first that $v \in (T - \lambda I)^{\dim V}$. Then by the definition of $G(\lambda, T)$, $v \in G(\lambda, T)$, as desired.

Now suppose that $v \in G(\lambda, T)$. Then $(T - \lambda I)^j v = 0$ for some positive integer j. Thus, $v \in \text{null}(T - \lambda I)^j$. We divide into two cases ($j < \dim V$ and $j \geq \dim V$). If $j < \dim V$, then by Theorem 8.1, $v \in \text{null}(T - \lambda I)^j \subset \text{null}(T - \lambda I)^{\dim V}$, as desired. On the other hand, if $j \geq \dim V$, then by Theorem 8.3 $v \in \text{null}(T - \lambda I)^j = \text{null}(T - \lambda I)^{\dim V}$, as desired. ∎

- We now prove an analogous result to Theorem 5.2 for generalized eigenvectors.

Theorem 8.6. *Let $T \in \mathcal{L}(V)$. Suppose $\lambda_1, \dots, \lambda_m$ are distinct eigenvalues of T and $v_1, \dots, v_m$ are corresponding generalized eigenvectors. Then $v_1, \dots, v_m$ is linearly independent.*

Proof. Suppose $a_1, \dots, a_m \in \mathbb{F}$ are numbers such that

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

We will prove that each $a_j = 0$ one at a time. Let's begin.

Let $j \in \{1, \dots, n\}$ be arbitrary, and let k be the largest nonnegative integer such that $(T - \lambda_j I)^k v_j \neq 0$. Let

$$w = (T - \lambda_j I)^k v_j$$

Then by the definition of k,

$$\begin{aligned}(T - \lambda_j I)w = (T - \lambda_j I)^{k+1} v_1 = 0 \\ Tw = \lambda_j w\end{aligned}$$

It follows that for any $\lambda \in \mathbb{F}$, $(T - \lambda I)w = (\lambda_j - \lambda)w$, which in turn implies that

$$(T - \lambda I)^n w = (\lambda_j - \lambda)^n w$$

for any $\lambda \in \mathbb{F}$ where $n = \dim V$. Thus, we have that

$$\begin{aligned}(T - \lambda_j I)^k \prod_{\substack{i=1 \\ i \neq j}}^m (T - \lambda_i I)^n 0 &= (T - \lambda_j I)^k \prod_{\substack{i=1 \\ i \neq j}}^m (T - \lambda_i I)^n (a_1 v_1 + \cdots + a_m v_m) \\ 0 &= a_1 (T - \lambda_j I)^k \prod_{\substack{i=1 \\ i \neq j}}^m (T - \lambda_i I)^n v_1 + \cdots + a_{j-1} (T - \lambda_j I)^k \prod_{\substack{i=1 \\ i \neq j}}^m (T - \lambda_i I)^n v_{j-1} \\ &\quad + a_j (T - \lambda_j I)^k \prod_{\substack{i=1 \\ i \neq j}}^m (T - \lambda_i I)^n v_j \\ &\quad + a_{j+1} (T - \lambda_j I)^k \prod_{\substack{i=1 \\ i \neq j}}^m (T - \lambda_i I)^n v_{j+1} + \cdots + a_m (T - \lambda_j I)^k \prod_{\substack{i=1 \\ i \neq j}}^m (T - \lambda_i I)^n v_m\end{aligned}$$

$$
\begin{aligned}
&= a_1(T-\lambda_jI)^k\left(\prod_{\substack{i=1\\ i\neq j,1}}^{m}(T-\lambda_iI)^n\right)(T-\lambda_1I)^nv_1\\
&\quad+\cdots+a_{j-1}(T-\lambda_jI)^k\left(\prod_{\substack{i=1\\ i\neq j,j-1}}^{m}(T-\lambda_iI)^n\right)(T-\lambda_{j-1}I)^nv_{j-1}\\
&\quad+a_j\left(\prod_{\substack{i=1\\ i\neq j}}^{m}(T-\lambda_iI)^n\right)(T-\lambda_jI)^kv_j\\
&\quad+a_{j+1}(T-\lambda_jI)^k\left(\prod_{\substack{i=1\\ i\neq j,j+1}}^{m}(T-\lambda_iI)^n\right)(T-\lambda_{j+1}I)^nv_{j+1}\\
&\quad+\cdots+a_m(T-\lambda_jI)^k\left(\prod_{\substack{i=1\\ i\neq j,m}}^{m}(T-\lambda_iI)^n\right)(T-\lambda_mI)^nv_m && \text{Theorem 5.4}\\
&= a_j\prod_{\substack{i=1\\ i\neq j}}^{m}(T-\lambda_iI)^nw && \text{Theorem 8.5}\\
&= a_j\prod_{\substack{i=1\\ i\neq j}}^{m}(\lambda_j-\lambda_i)^nw
\end{aligned}
$$

so $a_j=0$, as desired. ∎

- **Nilpotent** (operator): An operator T such that $T^j=0$ for some positive integer j.
- We now show that we never need to raise a nilpotent operator to a $j>\dim V$ to make it equal to zero.

Theorem 8.7. *Suppose $N\in\mathcal{L}(V)$ is nilpotent. Then $N^{\dim V}=0$.*

Proof. Since N is nilpotent, we know that there exists a nonnegative integer j such that

$$(N-0I)^jv=N^jv=0=0v$$

for any $v\in V$. Thus, $G(0,N)=V$. It follows by Theorem 8.5 that $V=G(0,N)=\operatorname{null}(N-0I)^{\dim V}=\operatorname{null}N^{\dim V}$. Consequently, for any $v\in V$, $N^{\dim V}v=0$, so $N^{\dim V}=0$, as desired. ∎

- We now show that if N is nilpotent, there exists a basis of V such that $\mathcal{M}(N)$ is more than half zeroes.

Theorem 8.8. *Suppose N is a nilpotent operator on V. Then there is a basis of V with respect to which the matrix of N has the form*

$$\begin{pmatrix}0 & & *\\ & \ddots & \\ 0 & & 0\end{pmatrix}$$

, i.e., where all entries on and below the diagonal are zeroes.

Proof. First choose a basis of $\operatorname{null}N$. Then extend this to a basis of $\operatorname{null}N^2$, then to a basis of $\operatorname{null}N^3$, on and on up until we have extended it to a basis $v_1,\dots,v_n$ of $\operatorname{null}N^{\dim V}$ (which, incidentally, will be a basis of V since $\operatorname{null} N^{\dim V} = V$ by Theorem 8.7). We will prove that $\mathcal{M}(N,(v_1,\dots,v_n))$ has the desired form.

Let k be the smallest positive integer such that $v_1 \in \operatorname{null} N^k$. Then $0 = N^k v_1 = N^{k-1} N v_1$, so $Nv_1 \in \operatorname{null} N^{k-1} = \{0\}$ by the condition on k. It follows that $Nv_1 = 0$, so since $v_1,\dots,v_n$ is linearly independent (as a basis), $\mathcal{M}(N,(v_1,\dots,v_n))_{\cdot,1} = \mathcal{M}(Nv_1)$ has only zero entries. Apply the same argument to any other vector in $\operatorname{null} N^k$, getting all zero columns for some number of columns. Having done this, move onto the first vector in the basis that is not in $\operatorname{null} N^k$. Let this vector be v_i. Then in a similar fashion to before, $Nv_i \in \operatorname{null} N^k$, so Nv_i is a linear combination of all vectors before v_i. Thus, all nonzero entries in $\mathcal{M}(()N,(v_1,\dots,v_n))_{\dots,i} = \mathcal{M}(()Nv_i)$ are above the diagonal. We continue in this fashion for the whole basis. ■

Exercises

1 Suppose $N \in \mathcal{L}(V)$ is nilpotent. Prove that 0 is the only eigenvalue of N.

Proof. Suppose for the sake of contradiction that $\lambda \neq 0$ is an eigenvalue of N with corresponding eigenvector v. Then

$$0 = 0v = N^{\dim V} v = \lambda^{\dim V} v \neq 0$$

a contradiction. ■

8.B Decomposition of an Operator

10/23:

- We are going to build up in this section to a proof that while not every operator's domain can be decomposed into eigenspaces, every operator's domain can be decomposed into generalized eigenspaces.
- We first show that the null and rance spaces of every polynomial of an operator T are invariant under T.

Theorem 8.9. *Suppose $T \in \mathcal{L}(V)$ and $p \in \mathcal{P}(\mathbb{F})$. Then $\operatorname{null} p(T)$ and $\operatorname{range} p(T)$ are invariant under T.*

Proof. To prove that $\operatorname{null} p(T)$ is invariant under T, it will suffice to show that for any $v \in \operatorname{null} p(T)$, $Tv \in \operatorname{null} p(T)$. Let $v \in \operatorname{null} p(T)$ be arbitrary. Then $(p(T))v = 0$. It follows that

$$\begin{aligned}
(p(T))(Tv) &= (p(T)T)v \\
&= (Tp(T))v && \text{Theorem 5.4} \\
&= T(p(T)v) \\
&= T(0) \\
&= 0 && \text{Theorem 3.2}
\end{aligned}$$

Therefore, $Tv \in \operatorname{null} p(T)$, as desired.

To prove that $\operatorname{range} p(T)$ is invariant under T, it will suffice to show that for any $v \in \operatorname{range} p(T)$, $Tv \in \operatorname{range} p(T)$. Let $v \in \operatorname{range} p(T)$ ve arbitrary. Then there exists $u \in V$ such that $p(T)u = v$. It follows that

$$\begin{aligned}
Tv &= T(p(T)u) \\
&= p(T)(Tu) && \text{Theorem 5.4}
\end{aligned}$$

Therefore, $Tv \in \operatorname{range} p(T)$, as desired. ■

- We now prove the main result we've been working up to. It shows that "every operator on a complex vector space can be thought of as composed of pieces, each of which is a nilpotent operator plus a scalar multiple of the identity" (Axler, 2015, p. 252).

Theorem 8.10. *Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T. Then*

(a) $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$.

Proof. We induct on $n = \dim V$. For the base case $n = 1$, since we know that there exists *an* eigenvalue of T (see Theorem 5.5) to which there can only correspond one set of eigenvector of T (because V is one-dimensional), we have that $V = G(\lambda_1, T)$, as desired. Now suppose using strong induction that we have proven the claim for all dimensions strictly less than n; we now seek to prove it for n. Let's begin.

By Theorem 5.5, T has an eigenvalue λ_1. Thus,

$$\begin{aligned} V &= \operatorname{null}(T - \lambda_1 I)^n \oplus \operatorname{range}(T - \lambda_1 I)^n && \text{Theorem 8.4} \\ &= G(\lambda_1, T) \oplus U && \text{Theorem 8.5} \end{aligned}$$

where we let $U = \operatorname{range}(T - \lambda_1 I)^n$. Under this definition, we have by Theorem 8.9 with $p(z) = (z - \lambda_1)^n$ that U is invariant under T. Additionally, since λ_1 is an eigenvalue of T (thus with some corresponding generalized eigenvector), $G(\lambda_1, T) \neq \{0\}$. Therefore, $\dim U < n$ by Theorem 3.22.

Consider $T|_U$. None of the generalized eigenvectors of $T|_U$ correspond to the eigenvalue λ_1, because all generalized eigenvectors of T corresponding to λ_1 are elements of $G(\lambda_1, T)$. Thus, the eigenvalues of $T|_U$ are exactly $\lambda_1, \dots, \lambda_m$.

Having established that $\dim U < n$ and $\lambda_2, \dots, \lambda_m$ are the distinct eigenvalues of $T|_U$, we have by the induction hypothesis that $U = G(\lambda_2, T|_U) \oplus \cdots + G(\lambda_m, T|_U)$. Consequently, all that's left is to show that $G(\lambda_k, T|_U) = G(\lambda_k, T)$ for each $k = 2, \dots, m$. Clearly, $G(\lambda_k, T|_U) \subset G(\lambda_k, T)$. In the other direction, suppose that $v \in G(\lambda_k, T)$. As an element of $V = G(\lambda_1, T) \oplus U$, we have that $v = v_1 + u$ where $v_1 \in G(\lambda_1, T)$ and $u \in U$. However, since v and v_1 correspond to different eigenvalues, we have by Theorem 8.6 that they are linearly independent. Thus, $v_1 = 0$. Therefore, $v = u \in U$, so $v \in G(\lambda_k, T|_U)$, as desired. ■

(b) *Each $G(\lambda_j, T)$ is invariant under T.*

Proof. By Theorem 8.5, $G(\lambda_j, T) = \operatorname{null}(T - \lambda_j I)^{\dim V}$. By Theorem 8.9, if $p(z) = (z - \lambda_j)^{\dim V}$, then $\operatorname{null} p(T) = \operatorname{null}(T - \lambda_j I)^{\dim V}$ is invariant under T. Therefore, $G(\lambda_j, T)$ is invariant under T, as desired. ■

(c) *Each $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent.*

Proof. Let $v \in G(\lambda_j, T)$. Then

$$\begin{aligned} ((T - \lambda_j I)|_{G(\lambda_j, T)})^{\dim V} v &= (T - \lambda_j I)^{\dim V} v \\ &= 0 && \text{Theorem 8.5} \end{aligned}$$

Thus, since there is a positive integer power of $(T - \lambda_j I)|_{G(\lambda_j, T)}$ such that $(T - \lambda_j I)|_{G(\lambda_j, T)} = 0$, $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent, as desired. ■

- We now show that while T may not have enough eigenvectors to form an eigenbasis, T always has enough generalized eigenvectors to form a basis.

Theorem 8.11. *Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then there is a basis of V consisting of generalized eigenvectors of T.*

Proof. Choose a basis of each $G(\lambda_j, T)$ proven to compose V in Theorem 8.10a. Concatenate these bases to form a basis of V consisting of generalized eigenvectors of V. ■

- **Multiplicity** (of an eigenvalue λ of $T \in \mathcal{L}(V)$): The dimension of $G(\lambda, T)$. *Also known as* **algebraic multiplicity**.

- We now prove an obvious consequence of the definition of multiplicity.

Theorem 8.12. *Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then the sum of the multiplicities of all the eigenvalues of T equals* $\dim V$.

Proof. We have from Theorem 8.10a that

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$$

Thus, by Theorem 3.22,

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

Therefore, by the definition of the multiplicity of an eigenvalue of T, the above equation proves the desired result. ■

- **Geometric multiplicity** (of an eigenvalue λ of $T \in \mathcal{L}(V)$): The dimension of $E(\lambda, T)$.
- **Block diagonal matrix**: 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 the other entries of the matrix equal 0.

- We now prove the results we have proven before, but in matrix form.

Theorem 8.13. *Suppose V is a complex vector space 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 is 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_j is a d_j-by-d_j upper-triangular matrix of the form

$$A_j = \begin{pmatrix} \lambda_j & & * \\ & \ddots & \\ 0 & & \lambda_j \end{pmatrix}$$

Proof. Let $j \in \{1, \dots, m\}$ be arbitrary. By Theorem 8.10c, $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent. Thus, by Theorem 8.8, we can choose a basis of $G(\lambda_j, T)$ (which will be of length d_j) such that $\mathcal{M}((T - \lambda_j I)|_{G(\lambda_j, T)})$ with respect to this basis has all zeroes on and below the diagonal. It follows that $\mathcal{M}(T|_{G(\lambda_j, T)}) = \mathcal{M}((T - \lambda_j I)|_{G(\lambda_j, T)}) + \mathcal{M}(\lambda_j I|_{G(\lambda_j, T)})$ will have the necessary form to be an A_j. Therefore, since concatenating the bases of each $G(\lambda_j, T)$ gives a basis of V by Theorem 8.10a, $\mathcal{M}(T)$ with respect to this basis will have the desired form. ■

- We now harness the power of some of our newer theorems to prove some further results about square roots.

Theorem 8.14. *Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then $I + N$ has a square root.*

Proof. Consider the Taylor series expansion of $\sqrt{I+N}$, as if I were the number 1 and N were some number x. We would have

$$\sqrt{I+N} = I + a_1 N + a_2 N^2 + \cdots$$

for some set of coefficients $a_1, a_2, \ldots$. Now while an infinite sum of operators cannot be an operator, this sum can: Since N is nilpotent, $N^m = 0$ for some positive integer m, so every term of degree $j \geq m$ is zero and the sum is finite. Indeed, for the sum above to be a square root of $I + N$, we need only require that

$$\begin{aligned} I + N &= \left(\sqrt{I+N}\right)^2 \\ &= \left(I + a_1 N + \cdots + a_{m-1} N^{m-1}\right)^2 \\ &= I + 2a_1 N + (2a_2 + a_1^2) N^2 + (2a_3 + 2a_1 a_2) N^3 + \cdots + cN^{m-1} \end{aligned}$$

where c stands in for a much more complicated coefficient in terms of $a_1, \ldots, a_{m-1}$. To do so, simply choose a_1 such that $2a_1 = 1$ (i.e., choose $a_1 = 1/2$), choose a_2 such that $2a_2 + a_1^2 = 0$ (i.e., choose $a_2 = -1/8$), choose a_3 such that $2a_3 + 2a_1 a_2 = 0$, and on and on. ■

Theorem 8.15. *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. Let $j \in \{1, \ldots, m\}$ be arbitrary. Since $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent (see Theorem 8.10c), we have that there exists a nilpotent operator $N_j \in \mathcal{L}(G(\lambda_j, T))$ such that $T|_{G(\lambda_j,T)} = \lambda_j I + N_j$. Additionally, we know that $\lambda_j \neq 0$: Since T is an operator on a finite-dimensional complex vector space, Theorem 5.7 implies that T has an upper-triangular matrix with respect to some basis of V; since T is invertible, Theorem 5.8 implies all of the diagonal entries of this matrix are nonzero; since Theorem 5.9 implies that the eigenvalues of T are exactly the diagonal entries of this matrix, we know that they are nonzero, i.e., in particular, $\lambda_j \neq 0$. Thus, since we can divide by λ_j, we have that

$$T|_{G(\lambda_j,T)} = \lambda_j \left(I + \frac{N_j}{\lambda_j}\right)$$

By Theorem 8.14, $I + N_j/\lambda_j$ has a square root. Naturally, this operator times a square root of λ_j is a square root R_j of $T|_{G(\lambda_j,T)}$. This combined with the fact that any $v \in V$ can be written uniquely in the form $u_1 + \cdots + u_m$ where $u_i \in G(\lambda_i, T)$ for each $i = 1, \ldots, m$ allows us to define the operator $R \in \mathcal{L}(V)$ by

$$Rv = R_1 u_1 + \cdots + R_m u_m$$

To prove that $R^2 = T$, it will suffice to show that $R^2 v = Tv$ for all $v \in V$. Let $v \in V$ be arbitrary. Then it follows from all of our definitions that

$$\begin{aligned} R^2 v &= R_1^2 u_1 + \cdots + R_m^2 u_m \\ &= \lambda_1 \left(I + \frac{N_1}{\lambda_1}\right) u_1 + \cdots + \lambda_m \left(I + \frac{N_m}{\lambda_m}\right) u_m \\ &= T|_{G(\lambda_1,T)} u_1 + \cdots + T|_{G(\lambda_m,T)} u_m \\ &= Tu_1 + \cdots + Tu_m \\ &= Tv \end{aligned}$$

as desired. ■

- Note that the techniques in this section can be adapted to prove that if V is a complex vector space and $T \in \mathcal{L}(V)$ is invertible, then T has a k^{th} root for every positive integer k.

Exercises

10/25: **9** Suppose A and B are block diagonal matrices of the form

$$A = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix} \qquad B = \begin{pmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_m \end{pmatrix}$$

where A_j has the same size as B_j for $j = 1, \dots, m$. Show that AB is a block diagonal matrix of the form

$$AB = \begin{pmatrix} A_1B_1 & & 0 \\ & \ddots & \\ 0 & & A_mB_m \end{pmatrix}$$

Proof. Suppose that A, B are $n \times n$ matrices. By the definition of matrix multiplication, if $1 \leq i, k \leq \dim A_1 = \dim B_1$ we have that

$$\begin{aligned} (AB)_{i,k} &= \sum_{r=1}^{n} A_{i,r}B_{r,k} \\ &= \sum_{r=1}^{\dim A_1} A_{i,r}B_{r,k} + \sum_{\dim A_1+1}^{n} 0 \cdot 0 \\ &= \sum_{r=1}^{\dim A_1} A_{1_{i,r}}B_{2_{r,k}} \\ &= (A_1B_1)_{i,k} \end{aligned}$$

We can do something similar for every other submatrix along the diagonal. ∎

8.C Characteristic and Minimal Polynomials

10/23:

- **Characteristic polynomial** (of $T \in \mathcal{L}(V)$, V complex): The polynomial

  $$(z - \lambda_1)^{d_1} \cdots (z - \lambda_m)^{d_m}$$

  where $\lambda_1, \dots, \lambda_m$ are the distinct eigenvalues of T with corresponding multiplicities $d_1, \dots, d_m$.

- We first show some elementary properties of the characteristic polynomial.

  Theorem 8.16. *Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then*

  (*a*) *The characteristic polynomial of T has degree* $\dim V$.

  Proof. This follows from the definition of the characteristic polynomial and Theorem 8.12. ∎

  (*b*) *The zeroes of the characteristic polynomial of T are the eigenvalues of T.*

  Proof. This follows from the definition of the characteristic polynomial. ∎

- The simple definition (as opposed to the determinant-based definition) of the characteristic polynomial given here affords a simple proof of the Complex Cayley-Hamilton Theorem.

  Theorem 8.17 (Complex Cayley-Hamilton Theorem[1])**.** *Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Let q denote the characteristic polynomial of T. Then $q(T) = 0$.*

[1]This result is named for English mathematician Arthur Cayley and Irish mathematician William Rowan Hamilton, both of whom found great success even before the completion of their undergraduate degrees.

Proof. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T. To prove that $q(T) = 0$, it will suffice to show that $q(T)v = 0$ for all $v \in V$. Let $v \in V$ be arbitrary. Then by Theorem 8.10a, $v = u_1 + \cdots + u_m$ where each $u_j \in G(\lambda_j, T)$. Additionally, Theorems 8.10c and 8.7 assert that each $(T - \lambda_j I)^{d_j}|_{G(\lambda_j,T)} = 0$. Therefore, we have that

$$q(T)v = \prod_{i=1}^{m} (T - \lambda_i I)^{d_i} (u_1 + \cdots + u_m) = 0$$

since after distributing the operator to each term in the sum, we can restrict the domain of each exponential to $G(\lambda_j, T)$ and commute the $(T - \lambda_j I)^{d_j}|_{G(\lambda_j,T)}$ term to be applied first (by Theorem 5.4). ■

- **Monic polynomial**: A polynomial whose highest-degree coefficient equals 1.
- We now prove that we can associate a unique **minimal polynomial** with each operator T.

Theorem 8.18. *Suppose $T \in \mathcal{L}(V)$. Then there is a unique monic polynomial p of smallest degree such that $p(T) = 0$.*

Proof. We first show that such a polynomial exists; then we prove its uniqueness. Let's begin.

Let $n = \dim V$. Consider the list of operators $I, T, T^2, \dots, T^{n^2} \in \mathcal{L}(V)$. Since $\dim \mathcal{L}(V) = n^2$ (see Theorem 3.16) and the length of this list is $n^2 + 1 > n^2$, Theorem 2.3 implies that this list is linearly dependent. Let m be the smallest positive integer such that

$$I, T, T^2, \dots, T^m$$

is linearly dependent.

By the Linear Dependence Lemma, one of the operators in the above list is a linear combination of the previous ones. By the choice of m, we know that this operator is T^m. Thus, there exist scalars $a_0, \dots, a_{m-1} \in \mathbb{F}$ such that

$$a_0 I + a_1 T + a_2 T^2 + \cdots + a_{m-1} T^{m-1} + T^m = 0$$

Therefore, if we consider the monic polynomial $p \in \mathcal{P}(\mathbb{F})$ defined by

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_{m-1} z^{m-1} + z^m$$

we know by the above that $p(T) = 0$.

Suppose for the sake of contradiction that there exists a monic polynomial $q \in \mathcal{P}(\mathbb{F})$ of degree less than or equal to m such that $q(T) = 0$. If $\deg q < m$, then $I, T, T^2, \dots, T^{\deg q}$ will be linearly independent, so $q(z) = 0$. But this implies the highest degree coefficient of q is not 1, so q is not monic, a contradiction. On the other hand, if $\deg q = m$, then we have that $(p - q)(T) = 0$ as well. However, since both p and z have a $1z^m$ term that cancels in $p - q$, $\deg(p - q) < m$. Thus, we can reach the same contradiction in the other case. ■

- **Minimal polynomial** (of T): The unique monic polynomial p of smallest degree such that $p(T) = 0$.
  - By the proof of Theorem 8.18, the degree of the minimal polynomial of each operator on V is at most $(\dim V)^2$.
  - By the Complex Cayley-Hamilton Theorem, the degree of the minimal polynomial of each operator on V *complex* is at most $\dim V$.
  - The minimal polynomial can be computed by considering a homogeneous system of equations

    $$a_0 \mathcal{M}(I) + a_1 \mathcal{M}(T) + \cdots + a_m \mathcal{M}(T)^m = 0$$

    with $(\dim V)^2$ equations in $a_0, \dots, a_m$ for successive values of m until a solution exists. This solution would give the coefficients of the minimal polynomial.

- We now characterize all polynomials that when applied to an operator give the 0 operator.

  Theorem 8.19. *Suppose $T \in \mathcal{L}(V)$ and $q \in \mathcal{P}(\mathbb{F})$. Then $q(T) = 0$ if and only if q is a polynomial multiple of the minimal polynomial of T.*

  Proof. Let p denote the minimal polynomial of T.

  Suppose first that $q(T) = 0$. By the Division Algorithm for Polynomials, we have that $q = ps + r$ where $\deg r < \deg p$. Thus,
  $$0 = q(T) = p(T)s(T) + r(T) = r(T)$$
  It follows that $r = 0$ (otherwise, since $\deg r < \deg p$, r divided by the coefficient of the highest-order term would be a monic polynomial that when applied to T of degree less than minimal polynomial, a contradiction). Therefore, q is a polynomial multiple of the minimal polynomial of T, as desired.

  Now suppose that q is a polynomial multiple of the minimal polynomial of T. The $q = ps$ for some $s \in \mathcal{P}(\mathbb{F})$. It follows that
  $$q(T) = p(T)s(T) = 0s(T) = 0$$
  as desired. ∎

- We can now apply our discussion of the minimal polynomial back to the characteristic polynomial.

  Theorem 8.20. *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.*

  Proof. Let q denote the characteristic polynomial of T. By the Complex Cayley-Hamilton Theorem, $q(T) = 0$. Therefore, by Theorem 8.19, $q(T)$ is a multiple of the minimal polynomial of T, as desired. ∎

- We now show that like the eigenvalues of T are the zeroes of the characteristic polynomial, they are the zeroes of the minimal polynomial[2].

  Theorem 8.21. *Let $T \in \mathcal{L}(V)$. Then the zeroes of the minimal polynomial of T are precisely the eigenvalues of T.*

  Proof. Let
  $$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_{m-1} z^{m-1} + z^m$$
  be the minimal polynomial of T. We will first show that every zero of p is an eigenvalue of T. Then, we will show that every eigenvalue of T is a zero of p. Let's begin.

  Suppose first that $\lambda \in \mathbb{F}$ satisfies $p(\lambda) = 0$. Thus, by Theorem 4.3, there exists a polynomial $q \in \mathcal{P}(\mathbb{F})$ such that $p(z) = (z - \lambda)q(z)$, where q must be a monic polynomial since p is one. Consequently, since $p(T) = 0$, we have that
  $$0 = (T - \lambda I)q(T)v$$
  for all $v \in V$. In particular, we must have $q(T)v \neq 0$ for some $v \in V$ (otherwise $q(T)$ with $\deg q < \deg p$ would be the minimal polynomial of T). Therefore, λ is an eigenvalue of T with corresponding eigenvector $q(T)v$ for this v, as desired.

  Now suppose that $\lambda \in \mathbb{F}$ is an eigenvalue of T. Then there exists a nonzero $v \in V$ such that $Tv = \lambda v$. It follows that $T^j v = \lambda^j v$ for every nonnegative integer j. Thus, we have that
  $$\begin{aligned}
  0 &= p(T)v \\
  &= (a_0 + a_1 T + a_2 T^2 + \cdots + a_{m-1} T^{m-1} + T^m)v \\
  &= (a_0 + a_1 \lambda + a_2 \lambda^2 + \cdots + a_{m-1} \lambda^{m-1} + \lambda^m)v \\
  &= p(\lambda)v
  \end{aligned}$$
  But since $v \neq 0$, we must have $p(\lambda) = 0$, as desired. ∎

- Axler (2015) gives some examples of how the previous results can be applied to tangible problems.

[2] It would appear then that the minimal polynomial can be written in the form $(z - \lambda_1) \cdots (z - \lambda_m)$.

8.D Jordan Form

- Theorem 8.13 got us to a pretty nice form for every operator T. We now build up to an even nicer one.
- We first show that every nilpotent operator has a corresponding basis consisting of certain powers of N applied to a select number of vectors.

Theorem 8.22. *Suppose $N \in \mathcal{L}(V)$ is nilpotent. Then there exist vectors $v_1, \dots, v_n \in V$ and nonnegative integers $m_1, \dots, m_n$ such that*

(*a*) $N^{m_1}v_1, \dots, Nv_1, v_1, \dots, N^{m_n}v_n, \dots, Nv_n, v_n$ *is a basis of V.*

(*b*) $N^{m_1+1}v_1 = \cdots = N^{m_n+1}v_n = 0$.

Proof. We induct on $n = \dim V$. For the base case $n = 1$, 0 is the only nilpotent operator on V. Thus, choose any nonzero v_1 and let $m_1 = 0$.

Now suppose using strong induction that we have proven the claim for all vector spaces of dimension less than n; we now seek to prove the claim for n. First off, note that N is not injective: If j is the smallest nonnegative integer such that of $N^j = 0$, then there exists $v \in V$ such that $N^{j-1}v \neq 0$; it follows that although $v \neq 2v$ and hence $N^{j-1}v \neq 2N^{j-1}v = N^{j-1}2v$, $N(N^{j-1}v) = 0$ and $N(N^{j-1}(2v)) = 2N(N^{j-1}v) = 2 \cdot 0 = 0$. It follows by Theorem 3.19 that N is not surjective. Thus, $\operatorname{range} N \neq V$, so we must have $\dim \operatorname{range} N < \dim V$.

Consider $N|_{\operatorname{range} N} \in \mathcal{L}(\operatorname{range} N)$, to which we can apply our inductive hypothesis by the previous result. Doing so, we find that there exist vectors $v_1, \dots, v_n \in \operatorname{range} N$ and nonnegative integers $m_1, \dots, m_n$ such that

$$N^{m_1}v_1, \dots, Nv_1, v_1, \dots, N^{m_n}v_n, \dots, Nv_n, v_n \tag{8.1}$$

is a basis of $\operatorname{range} N$ and

$$N^{m_1+1}v_1 = \cdots = N^{m_n+1}v_n = 0$$

Let $j \in \{1, \dots, n\}$ be arbitrary. Since $v_j \in \operatorname{range} N$, there exists $u_j \in V$ such that $Nu_j = v_j$. It follows that $N^{k+1}u_j = N^k v_j$ for each j and every nonnegative integer k. We now seek to prove that

$$N^{m_1+1}u_1, \dots, Nu_1, u_1, \dots, N^{m_n+1}u_n, \dots, Nu_n, u_n \tag{8.2}$$

is a linearly independent list of vectors in V. To do so, consider a linear combination of the vectors in List 8.2 that is equal to zero. Applying N to said linear combination gives a linear combination of List 8.1 that is equal to zero with related coefficients (since $Nu_j = v_j$, $N(N^k u_j) = N^k v_j$ for all $1 \leq k \leq m_j$, and $N(N^{m_j+1}u_j) = N^{m_j+1}v_j = 0$ for each $j = 1, \dots, n$). But since List 8.1 is linearly independent as a basis, all coefficients are zero with the possible exception of those of the vectors $N^{m_1+1}u_1, \dots, N^{m_n+1}u_n$, since those vectors go to zero as described above. However, once again, the linear independence of List 8.1, to which each of these vectors belongs in the form $N^{m_j}v_j$, implies that their coefficients are equal to zero as well. Thus, we have proven that List 8.2 is linearly independent, as desired.

Using Theorem 2.8, extend List 8.2 to a basis

$$N^{m_1+1}u_1, \dots, Nu_1, u_1, \dots, N^{m_n+1}u_n, \dots, Nu_n, u_n, w_1, \dots, w_p \tag{8.3}$$

of V. Let $j \in \{1, \dots, p\}$ be arbitrary. We know that $Nw_j \in \operatorname{range} N$. Thus, Nw_j is in the span of List 8.1. But since each vector in List 8.1 equals N applied to a vector in List 8.2, we have that $Nw_j = Nx_j$ for some x_j in the span of List 8.2. Let

$$u_{n+j} = w_j - x_j$$

for each $j = 1, \dots, n$. Then

$$N^{m_1+1}u_1, \dots, Nu_1, u_1, \dots, N^{m_n+1}u_n, \dots, Nu_n, u_n, u_{n+1}, \dots, u_{n+p}$$

spans V since its span contains each x_j (in the span of the List 8.2 part) and each u_{n+j} and hence each w_j (and because List 8.3 spans V). Thus, since the above spanning list has the same length as the basis of V in List 8.3, Theorem 2.13 implies that it is a basis of V. It clearly has the required form (choose $m_j = 0$ for $j = n+1, \ldots, n+p$). Additionally, we have $Nu_{n+j} = Nw_j - Nx_j = 0$ for each $j = 1, \ldots, n$, verifying part (b). ∎

- **Jordan basis** (for $T \in \mathcal{L}(V)$): A basis of V such that with respect to this basis, T has a block diagonal matrix

$$\begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_p \end{pmatrix}$$

where each A_j is an upper-triangular matrix of the form

$$A_j = \begin{pmatrix} \lambda_j & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_j \end{pmatrix}$$

- We are now ready for the main result.

Theorem 8.23 (Jordan Form[3]). *Suppose V is a complex vector space. If $T \in \mathcal{L}(V)$, then there is a basis of V that is a Jordan basis for T.*

Proof. We divide into two cases (T is nilpotent and T is not nilpotent), using one to prove the other. Suppose first that T be a nilpotent operator on V. Consider the vectors $v_1, \ldots, v_n \in V$ associated with it by Theorem 8.22. Let $j \in \{1, \ldots, n\}$ be arbitrary. Notice that N sends the first vector in the list $N^{m_j}v_j, \ldots, Nv_j, v_j$ to 0 and every other vector to the previous one in the list. Thus, applying this observation to all j, we realize that the matrix of N with respect to the basis given by Theorem 8.22 is block diagonal with each matrix on the diagonal having the form

$$\begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix}$$

Combining this result with Exercise 8.A.1 proves that the desired result holds for nilpotent operators.

Now suppose that T is not a nilpotent operator. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. By Theorem 8.10c, each $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent. Thus, by the proof of the first case, there exists a Jordan basis for each $(T - \lambda_j I)|_{G(\lambda_j, T)}$. It follows since $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$ (see Theorem 8.10a) that concatenating all of these bases gives a basis of V that is a Jordan basis for T. ∎

[3]This result is named for French mathematician Camille Jordan, who published the first proof of this theorem.

Chapter 9

Operators on Real Vector Spaces

9.A Complexification

10/24:

- **Complexification** (of V): The set $V \times V$, where V is a real vector space. *Denoted by* $\boldsymbol{V_{\mathbb{C}}}$.
  - The complexification of V allows us to embed a real vector space in a complex vector space so that our results concerning operators on complex vector spaces can be translated into information about operators on real vector spaces.
- An element of $V_{\mathbb{C}}$ is an ordered pair (u, v), where $u, v \in V$, but we will 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 $u_1, v_1, u_2, v_2 \in V$.
- 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 $u, v \in V$.
- Thus, we can prove that $V \times V$ is a vector space.
- If we identify $u \in V$ with $u + 0i \in V_{\mathbb{C}}$, then we can think of V as a subset of $V_{\mathbb{C}}$.
  - Basically, the construction of $V_{\mathbb{C}}$ from V generalizes the construction of $\mathbb{C}^n$ from $\mathbb{R}^n$.
- Many things transfer nicely from V to $V_{\mathbb{C}}$, as exemplified by the following.

  Theorem 9.1. *Suppose V is a real vector space.*

  (a) *If $v_1, \dots, v_n$ is a basis of V, then $v_1, \dots, v_n$ is a basis of $V_{\mathbb{C}}$.*

  Proof. Let $v_1, \dots, v_n$ be a basis of V.
  To prove that $v_1, \dots, v_n$ spans $V_{\mathbb{C}}$, we will prove an inclusion in both directions. Clearly, $\operatorname{span}(v_1, \dots, v_n) \subset V_{\mathbb{C}}$. In the other direction, let $u + iv \in V_{\mathbb{C}}$ be arbitrary. Since $u, v \in V$,

$$u = \alpha_1 v_1 + \cdots + \alpha_n v_n \qquad\qquad v = \beta_1 v_1 + \cdots + \beta_n v_n$$

  for some $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n \in \mathbb{R}$. It follows that

$$\begin{aligned} u + iv &= \alpha_1 v_1 + \cdots + \alpha_n v_n + i(\beta_1 v_1 + \cdots + \beta_n v_n) \\ &= (\alpha_1 + i\beta_1)v_1 + \cdots + (\alpha_n + i\beta_n)v_n \end{aligned}$$

so $u + iv \in \operatorname{span}(v_1, \dots, v_n)$, as desired.

To prove that $v_1, \dots, v_n$ is linearly independent, suppose $\lambda_1, \dots, \lambda_n \in C$ make

$$\lambda_1 v_1 + \cdots + \lambda_n v_n = 0$$

Then naturally

$$(\operatorname{Re}\lambda_1)v_1 + \cdots + (\operatorname{Re}\lambda_n)v_n = 0 \qquad (\operatorname{Im}\lambda_1)v_1 + \cdots + (\operatorname{Im}\lambda_n)v_n = 0$$

so we must have that $\operatorname{Re}\lambda_j = \operatorname{Im}\lambda_j = \lambda_j = 0$ for each $j = 1, \dots, n$ since $v_1, \dots, v_n$ is linearly independent in V. ■

(*b*) *The dimension of $V_{\mathbb{C}}$ equals the dimension of V.*

Proof. This follows from part (a) by the definition of dimension. ■

- **Complexification** (of $T \in \mathcal{L}(V)$): The operator $T_{\mathbb{C}} \in \mathcal{L}(V_{\mathbb{C}})$ defined by

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

for all $u, v \in V$ a real vector space.

  - Note that technically, we must prove that $T_{\mathbb{C}}$ is *actually* in $\mathcal{L}(V_{\mathbb{C}})$ as defined.

- If V is a real vector space with basis $v_1, \dots, v_n$ and $T \in \mathcal{L}(V)$, then $\mathcal{M}(T, (v_1, \dots, v_n)) = \mathcal{M}(T_{\mathbb{C}}, (v_1, \dots, v_n))$.

  - The proof of this claim follows immediately from the definitions.

- We now apply complexification to answer a question about invariant subspaces.

Theorem 9.2. *Every operator on a nonzero finite-dimensional vector space has an invariant subspace of dimension 1 or 2.*

Proof. Let V be a nonzero finite-dimensional vector space, and let $T \in \mathcal{L}(V)$. We divide into two cases (V is complex and V is real).

If V is complex, then by Theorem 5.5, T has an eigenvalue and hence a corresponding eigenvector v. Thus, T has a 1-dimensional invariant subspace (namely, $\operatorname{span}(v)$).

If V is real, then by Theorem 5.5, $T_{\mathbb{C}}$ has an eigenvalue $a + bi$ and a corresponding eigenvector $u + iv$. It follows that

$$Tu + iTv = T_{\mathbb{C}}(u + iv) = (a + ib)(u + iv) = (au - bv) + (av + bu)i$$

i.e., that

$$Tu = au - bv \qquad Tv = av + bu$$

The above two equations prove that $\operatorname{span}(u, v)$ is an invariant subspace of V under T of dimension ≤ 2. ■

- Relating the minimal polynomials of $T_{\mathbb{C}}$ and T.

Theorem 9.3. *Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Then the minimal polynomial of $T_{\mathbb{C}}$ equals the minimal polynomial of T.*

Proof. Let $p \in \mathcal{P}(\mathbb{R})$ be the minimal polynomial of T. To prove that p is the minimal polynomial of $T_{\mathbb{C}}$, it will suffice to show that $p(T_{\mathbb{C}}) = 0$ and that if $q \in \mathcal{P}(\mathbb{C})$ is a monic polynomial such that $q(T_{\mathbb{C}}) = 0$, then $\deg q \geq \deg p$ (see Theorem 8.18 for the second claim). Let's begin.

For the first part, since $(T_{\mathbb{C}})^n(u + iv) = T^n u + iT^n v$ by the definition of $T_{\mathbb{C}}$, we have that $p(T_{\mathbb{C}}) = (p(T))_{\mathbb{C}} = 0_{\mathbb{C}} = 0$, as desired.

For the second part, suppose $q \in \mathcal{P}(\mathbb{C})$ is a monic polynomial such that $q(T_{\mathbb{C}}) = 0$. Then $(q(T_{\mathbb{C}}))u = 0$ for all $u = u + 0i \in V$. It follows that if $r \in \mathcal{P}(\mathbb{R})$ is the polynomial with j^{th} coefficient equal to the real part of the j^{th} coefficient of q, then $r(T) = 0$. Therefore, $\deg q = \deg r \geq \deg p$, as desired. ■

- An interesting corollary to the previous result is that the coefficients of the minimal polynomial of $T_{\mathbb{C}}$ are real.
- We now show that the *real* eigenvalues of the complexification of T are exactly the eigenvalues of T.

Theorem 9.4. *Suppose V is a real vector space, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbb{R}$. Then λ is an eigenvalue of $T_{\mathbb{C}}$ if and only if λ is an eigenvalue of T.*

Proof[1]. Let $p \in \mathcal{P}(\mathbb{C})$ be the minimal polynomial of $T_{\mathbb{C}}$. Then we have that

$$\begin{aligned} \lambda \text{ is a real eigenvalue of } T_{\mathbb{C}} &\iff \lambda \text{ is a real zero of } p(T_{\mathbb{C}}) && \text{Theorem 8.21} \\ &\iff \lambda \text{ is a zero of } p(T) && \text{Theorem 9.3} \\ &\iff \lambda \text{ is an eigenvalue of } T && \text{Theorem 8.21} \end{aligned}$$

as desired. ■

- We now show that $T_{\mathbb{C}}$ treats $\lambda, \bar{\lambda}$ the same way.

Theorem 9.5. *Suppose V is a real vector space, $T \in \mathcal{L}(V)$, $\lambda \in \mathbb{C}$, j is a nonnegative integer, and $u, v \in V$. Then $(T_{\mathbb{C}} - \lambda I)^j(u + iv) = 0$ if and only if $(T_{\mathbb{C}} - \bar{\lambda} I)^j(u - iv) = 0$.*

Proof. We induct on j.

For the base case $j = 0$, suppose first that $(T_{\mathbb{C}} - \lambda I)^0(u + iv) = u + iv = 0$. Then $u, v = 0$. It follows that $T_{\mathbb{C}} - \bar{\lambda} I)^0(u - iv) = u - iv = 0$, as desired. The proof is symmetric in the reverse direction.

Now suppose inductively that we have proven the claim for $j-1$; we now seek to prove it for j. Suppose first that $(T_{\mathbb{C}} - \lambda I)^j(u + iv) = 0$. Then

$$\begin{aligned} 0 &= (T_{\mathbb{C}} - \lambda I)^{j-1}((T_{\mathbb{C}} - \lambda I)(u + iv)) \\ &= (T_{\mathbb{C}} - \lambda I)^{j-1}((Tu - au + bv) + i(Tv - av - bu)) \end{aligned}$$

It follows by the inductive hypothesis that

$$\begin{aligned} 0 &= (T_{\mathbb{C}} - \lambda I)^{j-1}((Tu - au + bv) - i(Tv - av - bu)) \\ &= (T_{\mathbb{C}} - \lambda I)^{j-1}((T_{\mathbb{C}} - \lambda I)(u - iv)) \\ &= (T_{\mathbb{C}} - \bar{\lambda} I)^j(u - iv) \end{aligned}$$

as desired. The proof is symmetric in the other direction. ■

- We can now prove that having one complex number be an eigenvalue of $T_{\mathbb{C}}$ necessitates that its complex conjugate is an eigenvalue of $T_{\mathbb{C}}$.

Theorem 9.6. *Suppose V is a real vector space, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbb{C}$. Then λ is an eigenvalue of $T_{\mathbb{C}}$ if and only if $\bar{\lambda}$ is an eigenvalue of $T_{\mathbb{C}}$.*

Proof. Take $j = 1$ in Theorem 9.5. ■

- Since a real operator can naturally only have real eigenvalues, "when mathematicians sometimes informally mention the complex eigenvalues of an operator on a real vector space, what they have in mind is the eigenvalues of the complexification of the operator" (Axler, 2015, p. 281)..
- We now prove that the multiplicities of complex conjugate eigenvalues coincide.

Theorem 9.7. *Suppose V is a real vector space, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbb{C}$ is an eigenvalue of $T_{\mathbb{C}}$. Then the multiplicity of λ as an eigenvalue of $T_{\mathbb{C}}$ equals the multiplicity of $\bar{\lambda}$ as an eigenvalue of $T_{\mathbb{C}}$.*

Proof. Suppose $u_1 + iv_1, \dots, u_m + iv_m$ is a basis of the generalized eigenspace $G(\lambda, T_{\mathbb{C}})$. Then with the help of Theorem 9.5, we can easily show that $u_1 - iv_1, \dots, u_m - iv_m$ is a basis of the generalized eigenspace $G(\bar{\lambda}, T_{\mathbb{C}})$. Therefore, the multiplicities coincide at m. ∎

- Although there exist operators on $\mathbb{R}^2$ (for example) with no eigenvalues, this is not true for every real vector space.

Theorem 9.8. *Every operator on an odd-dimensional real vector space has an eigenvalue.*

Proof. Suppose V is a real vector space with odd dimension, and let $T \in \mathcal{L}(V)$. Since every complex eigenvalue of $T_{\mathbb{C}}$ comes paired with its conjugate (see Theorem 9.6) and the members of each conjugate pair have the same multiplicity (see Theorem 9.7), the sum of multiplicities of the complex eigenvalues will be an even number. However, by Theorem 8.12, the sum of all of the multiplicities (counting the complex *and* real) of the eigenvalues of $T_{\mathbb{C}}$ will equal $\dim V_{\mathbb{C}}$, an odd number. Thus, there must be at least one additional eigenvalue λ of $T_{\mathbb{C}}$ that is not complex, i.e., is real. It follows by Theorem 9.4 that λ is an eigenvalue of T, as desired. ∎

- We now build up to defining the characteristic polynomial for real operators.

Theorem 9.9. *Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Then the coefficients of the characteristic polynomial of $T_{\mathbb{C}}$ are all real.*

Proof. Suppose λ is a nonreal eigenvalue of $T_{\mathbb{C}}$ with multiplicity m. Then by Theorems 9.6 and 9.7, $\bar{\lambda}$ is also an eigenvalue of $T_{\mathbb{C}}$ with multiplicity m. Thus, the characteristic polynomial of $T_{\mathbb{C}}$ includes the term

$$(z-\lambda)^m (z-\bar{\lambda})^m = (z^2 - 2(\operatorname{Re}\lambda)z + |\lambda|^2)^m$$

which has only real coefficients.

Since the characteristic polynomial of $T_{\mathbb{C}}$ is the product of terms with the above form and terms of the form $(z-t)^d$, where t is a real eigenvalue of $T_{\mathbb{C}}$ with multiplicity d, the coefficients of the characteristic polynomial of $T_{\mathbb{C}}$ are all real. ∎

- The above result allows for the following definition.
- **Characteristic polynomial** (of $T \in \mathcal{L}(V)$, V real): The characteristic polynomial of $T_{\mathbb{C}}$.
- Properties of the characteristic polynomial.

Theorem 9.10. *Suppose V is a real vector space and $T \in \mathcal{L}(V)$. Then*

(a) *The coefficients of the characteristic polynomial of T are all real.*

Proof. See Theorem 9.9. ∎

(b) *The characteristic polynomial of T has degree* $\dim V$*.*

Proof. See Theorem 8.16a. ∎

(c) *The eigenvalues of T are precisely the real zeroes of the characteristic polynomial of T.*

Proof. By Theorem 9.3, the real zeroes of the characteristic polynomial of T are the real zeroes of the characteristic polynomial of $T_{\mathbb{C}}$. These are, in turn, the real eigenvalues of $T_{\mathbb{C}}$ (by Theorem 8.16b). These, lastly, are in turn the eigenvalues of T (by Theorem 9.4). ∎

- We can now prove the complete Cayley-Hamilton Theorem.

Theorem 9.11 (Cayley-Hamilton Theorem)**.** *Suppose $T \in \mathcal{L}(V)$. Let q denote the characteristic polynomial of T. Then $q(T) = 0$.*

Proof. We divide into two cases (V is complex and V is real).

If V is complex, then apply the Complex Cayley-Hamilton Theorem.

If V is real, then by the Complex Cayley-Hamilton Theorem, $q(T_{\mathbb{C}}) = 0$. It follows by the definition of the characteristic polynomial of T that $q(T) = q(T_{\mathbb{C}}) = 0$, as desired. ■

- We now extend one last result from the complex to the real case.

Theorem 9.12. *Suppose $T \in \mathcal{L}(V)$. Then*

(*a*) *The degree of the minimal polynomial of T is at most* $\dim V$.

Proof. Let $p \in \mathcal{P}(\mathbb{R})$ be the minimal polynomial of T, and let $q \in \mathcal{P}(\mathbb{R})$ be the characteristic polynomial of T. By Theorem 9.10b, $\deg q = \dim V$. By the Cayley-Hamilton Theorem, $q(T) = 0$. Thus, by the definition of the minimal polynomial

$$\dim p \leq \dim q = \dim V$$

as desired. ■

(*b*) *The characteristic polynomial of T is a polynomial multiple of the minimal polynomial of T.*

Proof. Let $p \in \mathcal{P}(\mathbb{R})$ be the minimal polynomial of T, and let $q \in \mathcal{P}(\mathbb{R})$ be the characteristic polynomial of T. By the Complex Cayley-Hamilton Theorem, $q(T_{\mathbb{C}}) = 0$. Thus, by Theorem 8.19 $q(T) = q(T_{\mathbb{C}})$ is a polynomial multiple of $p(T_{\mathbb{C}}) = p(T)$ (where the last equality follows from Theorem 9.3). ■

9.B Operators on Real Inner Product Spaces

10/25:

- In this section, we characterize normal operators and isometries on real inner product spaces.
- First off, we describe normal but not self-adjoint operators on 2-dimensional real inner product spaces.

Theorem 9.13. *Suppose V is a 2-dimensional real inner product space and $T \in \mathcal{L}(V)$. Then the following are equivalent.*

(*a*) *T is normal but not self-adjoint.*

(*b*) *The matrix of T with respect to every orthonormal basis of V has the form*

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

with $b \neq 0$.

(*c*) *The matrix of T with respect to some orthonormal basis of V has the form*

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

with $b > 0$.

Proof. We will show that (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (a). Let's begin.

First, suppose that T is normal but not self-adjoint, and let e_1, e_2 be an arbitrary orthonormal basis of V. We have from the definitions that

$$\mathcal{M}(T, (e_1, e_2)) = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

for some $a, b, c, d \in \mathbb{R}$. We now look to express some of these variable in terms of others using known constraints on normal but not self-adjoint matrices. Since T is normal, Theorem 7.9 implies that $\|Te_1\|^2 = \|T^*e_1\|^2$. Thus, we have that

$$\begin{aligned} a^2 + b^2 = \|Te_1\|^2 = \|T^*e_1\|^2 &= a^2 + c^2 \\ b^2 &= c^2 \\ c &= \pm b \end{aligned}$$

Since T is not self-adjoint, we must choose $c = -b$ instead of $c = b$. Thus, we have that

$$\mathcal{M}(T, (e_1, e_2)) = \begin{pmatrix} a & -b \\ b & d \end{pmatrix}$$

Additionally, since T is normal, we have by definition that

$$\begin{aligned} TT^* &= T^*T \\ \begin{pmatrix} a & -b \\ b & d \end{pmatrix} \begin{pmatrix} a & b \\ -b & d \end{pmatrix} &= \begin{pmatrix} a & b \\ -b & d \end{pmatrix} \begin{pmatrix} a & -b \\ b & d \end{pmatrix} \\ \begin{pmatrix} a^2 + b^2 & ab - bd \\ ab - bd & b^2 + d^2 \end{pmatrix} &= \begin{pmatrix} a^2 + b^2 & -ab + bd \\ -ab + bd & b^2 + d^2 \end{pmatrix} \end{aligned}$$

Thus, we have that $ab = bd$. But since $b \neq 0$ (otherwise, T would be self-adjoint), we must have $d = a$, as desired.

Second, suppose that (b) holds. Since $b \neq 0$, either $b > 0$ or $b < 0$ in $\mathcal{M}(T)$ with respect to any orthonormal basis of V. We now divide into two cases. If $b > 0$, then we are done. On the other hand, if $b < 0$ in $\mathcal{M}(T, (e_1, e_2))$, we will have $b > 0$ in $\mathcal{M}(T, (e_1, -e_2))$, as desired.

Third, suppose that (c) holds. Since $b > 0$ in $\mathcal{M}(T)$ with respect to some orthonormal basis, the upper right and lower left entries in $\mathcal{M}(T)$ are distinct. Thus, T is not self-adjoint. However, we can use matrix multiplication to verify that the matrices of TT^* and T^*T are equal with respect to this basis, proving that T is normal, as desired. ■

- We now prove a lemma to our main description of normal operators.

Theorem 9.14. *Suppose V is an inner product space, $T \in \mathcal{L}(V)$ is normal, and U is a subspace of V that is invariant under T. Then*

(a) *$U^\perp$ is invariant under T.*

Proof. Let $e_1, \dots, e_m$ be an orthonormal basis of U. Extend it to an orthonormal basis $e_1, \dots, e_m, f_1, \dots, f_m$ of V (see Theorem 6.15). Since U is invariant under T each Te_j is a linear combination of $e_1, \dots, e_m$. Thus, $\mathcal{M}(T)$ is of the following form.

$$\mathcal{M}(T) = \begin{array}{c} \\ e_1 \\ \vdots \\ e_m \\ f_1 \\ \vdots \\ f_m \end{array} \begin{array}{c} \begin{array}{cccccc} e_1 & \cdots & e_m & f_1 & \cdots & f_m \end{array} \\ \left(\begin{array}{cc} A & B \\ 0 & C \end{array} \right) \end{array}$$

For each $j = 1, \dots, m$, Theorem 6.9 asserts that $\|Te_j\|^2$ is the sum of the squares of the absolute values of the entries in the j^{th} column of A. Similarly, for each $j = 1, \dots, m$, Theorem 6.9 asserts that $\|T^*e_j\|^2$ is the sum of the squares of the absolute values of the entries in the j^{th} columns of A and B. But since these two values are equal by Theorem 7.9, we must have that all of the values in the j^{th} column of B are 0 for each $j = 1, \dots, n$, i.e., that $B = 0$. Thus, $Tf_k \in \text{span}(f_1, \dots, f_n)$ for each $k = 1, \dots, n$. It follows since $f_1, \dots, f_n$ is a basis of $U^\perp$ that $Tv \in U^\perp$ for any $v \in U^\perp$, implying that $U^\perp$ is invariant under T, as desired. ■

(b) *U is invariant under T^*.*

Proof. By part (a), $\mathcal{M}(T^*)$ is of the same form as $\mathcal{M}(T)$. Thus, by a similar argument, T^* is invariant under U. ■

(c) $(T|_U)^* = (T^*)|_U$.

Proof. For every $u, v \in U$, we have that

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

Therefore, $(T|_U)^* = (T^*)|_U$, as desired. ■

(d) *$T|_U \in \mathcal{L}(U)$ and $T|_{U^\perp} \in \mathcal{L}(U^\perp)$ are normal operators.*

Proof. We have from the above results that

$$(T|_U)(T|_U)^* = T|_U(T^*)|_U = (TT^*)|_U = (T^*T)|_U = (T^*)|_U T|_U = (T|_U)^*(T|_U)$$

and similarly for $U^\perp$. ■

- Note that if an operator has a block diagonal matrix with respect to some basis, then the entry in any 1×1 block on the diagonal of this matrix is an eigenvalue of T.
- We now prove that all normal operators on real inner product spaces have block-diagonal matrices with blocks of size at most 2×2.

Theorem 9.15. *Suppose V is a real inner product space and $T \in \mathcal{L}(V)$. Then the following are equivalent:*

(a) *T is normal.*

(b) *There is an orthonormal basis of V with respect to which T has a block diagonal matrix such that each block is a 1×1 matrix or a 2×2 matrix of the form*

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

with $b > 0$.

Proof. Suppose first that T is normal. We induct on $n = \dim V$. For the base case $n = 1$, T will trivially have a block diagonal matrix with a single 1×1 block for any basis of V, in particular any orthonormal one we can pick by Theorem 6.14. For the base case $n = 2$, we divide into two cases (T is self-adjoint and T is not self-adjoint). If T is self-adjoint, then by the Real Spectral Theorem, there is an orthonormal basis of V with respect to which T has a diagonal matrix (equivalently, a block diagonal matrix where each [the one] block is a 1×1 matrix), as desired. If T is not self-adjoint, then since T is also normal by hypothesis, Theorem 9.13 asserts that there exists an orthonormal basis of V with respect to which T has a matrix of the form $\left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$ with $b > 0$ (equivalently, a block-diagonal matrix where each [the one] block is a 2×2 matrix of the desired form), as desired.

Now suppose using strong induction that we have proven the claim for $n - 1$; we now seek to prove it for n. By Theorem 9.2, T has an invariant subspace U of dimension 1 or 2. We now divide into two cases. If $\dim U = 1$, choose $u \in U$ such that $\|u\| = 1$; this vector forms an orthonormal basis of U. If $\dim U = 2$, then by Theorem 9.14, $T|_U \in \mathcal{L}(U)$ is normal. Additionally, $T|_U$ is not self-adjoint (otherwise, Theorem 7.14 would imply that T has an eigenvalue and hence a corresponding eigenvector, making $\dim U = 1$). Thus, by Theorem 9.13, we can choose an orthonormal basis of U with respect to which $\mathcal{M}(T|_U)$ has the required form. Either way, we now have $\dim U^\perp < \dim V$. This combined with the facts that $U^\perp$ is invariant under T and $T|_{U^\perp}$ is normal (see Theorem 9.14) allows us to apply our induction hypothesis. Doing so reveals that there is an orthonormal basis of $U^\perp$ with respect to which

$\mathcal{M}(T|_{U^\perp})$ has the desired form. Concatenating this basis to the previously found basis of U gives an orthonormal basis of V with respect to which the matrix of T has the desired form overall.

Now suppose that (b) holds. Then using Exercise 8.B.9, to show that $TT^* = T^*T$, it will suffice to show that all submatrices along the diagonal commute, too. The 1×1 submatrices obviously commute, and the 2×2 ones commute since

$$\begin{aligned}
T_jT_j^* &= \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \\
&= \begin{pmatrix} a^2+b^2 & 0 \\ 0 & a^2+b^2 \end{pmatrix} \\
&= \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \\
&= T_j^*T_j
\end{aligned}$$

as desired. ∎

10/27: • We now characterize isometries on real inner product spaces.

Theorem 9.16. *Suppose V is a real inner product space and $S \in \mathcal{L}(V)$. Then the following are equivalent.*

(a) *S is an isometry.*

(b) *There is an orthonormal basis of V with respect to which S has a block diagonal matrix such that each block on the diagonal is a 1×1 matrix containing 1 or -1 or is a 2×2 matrix of the form*

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

with $\theta \in (0, \pi)$.

Proof. Suppose first that S is an isometry. Then by Theorem 7.19, $S^*S = I = SS^*$, so S is normal. Thus, by Theorem 9.15, there is an orthonormal basis of V with respect to which S has a block diagonal matrix such that each block is a 1×1 matrix or a 2×2 matrix of the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

with $b > 0$.

Suppose λ is the sole entry in a 1×1 block of such a matrix. Then there is a basis vector e_j such that $Se_j = \lambda e_j$. But since S is an isometry,

$$1 = \|e_j\| = \|Se_j\| = \|\lambda e_j\| = |\lambda|$$

Since $1, -1$ are the only real numbers with absolute value equal to 1, we have $\lambda = 1$ or $\lambda = -1$, as desired.

Consider a 2×2 matrix of the above form along the diagonal of S. It follows that there are basis vectors e_j, e_{j+1} such that

$$Se_j = ae_j + be_{j+1}$$

Thus since S is an isometry and by Theorem 6.9, we have that

$$1 = \|e_j\|^2 = \|Se_j\|^2 = a^2 + b^2$$

This combined with the condition that $b > 0$ implies that there exists a number $\theta \in (0, \pi)$ such that $a = \cos\theta$ and $b = \sin\theta$. Thus, the 2×2 diagonal entry has the desired form.

Now suppose that (b) holds. Then there is a direct sum decomposition

$$V = U_1 \oplus \cdots \oplus U_m$$

where each U_j is a subspace of V of dimension 1 or 2 such that any two vectors belonging to distinct U's are orthogonal and each $S|_{U_j}$ is an isometry under which U_j is invariant. Now let $v \in V$ be arbitrary. It follows that

$$\begin{aligned}
\|Sv\|^2 &= \|Su_1 + \cdots + Su_m\|^2 \\
&= \|Su_1\|^2 + \cdots + \|Su_m\|^2 && \text{Theorem 6.9} \\
&= \|u_1\|^2 + \cdots + \|u_m\|^2 \\
&= \|v\|^2
\end{aligned}$$

so S is an isometry, as desired. ■

Chapter 10

Trace and Determinant

10.A Trace

10/27:

- To study the trace and determinant, we'll need to know how $\mathcal{M}(T, (v_1, \dots, v_n))$ (for $T \in \mathcal{L}(V)$) changes as $v_1, \dots, v_n$ changes.
- ***n*-by-*n* identity matrix**: The matrix of the identity operator $I \in \mathcal{L}(V)$. *Denoted by* $\boldsymbol{I}$. *Given by*

$$\begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$$

  - $\mathcal{M}(I)$ is the same with respect to every basis of V.
- **Invertible** (matrix A): A square matrix A for which there exists a square matrix B of identical size such that $AB = BA = I$. *Also known as* **nonsingular**.
- **Inverse** (of an invertible matrix A): The unique matrix B in the above definition. *Denoted by* $\boldsymbol{A^{-1}}$.
  - The "unique" part of this definition follows from a proof symmetric to that of Theorem 3.12.
- **Singular** (matrix A): A matrix A that is not invertible. *Also known as* **noninvertible**.
- The following result is connected to Theorem 3.11.

Theorem 10.1. *Suppose $u_1, \dots, u_n$, $v_1, \dots, v_n$, and $w_1, \dots, w_n$ are all bases of V. Suppose $S, T \in \mathcal{L}(V)$. Then*

$$\mathcal{M}(ST, (u_1, \dots, u_n), (w_1, \dots, w_n)) = \mathcal{M}(S, (v_1, \dots, v_n), (w_1, \dots, w_n))\mathcal{M}(T, (u_1, \dots, u_n), (v_1, \dots, v_n))$$

- We now discuss the matrix of the identity operator with respect to two bases.

Theorem 10.2. *Suppose $u_1, \dots, u_n$ and $v_1, \dots, v_n$ are bases of V. Then the matrices*

$$\mathcal{M}(I, (u_1, \dots, u_n), (v_1, \dots, v_n)) \qquad \mathcal{M}(I, (v_1, \dots, v_n), (u_1, \dots, u_n))$$

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

Proof. It follows from Theorem 10.1 that

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

and

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

as desired. ■

- It follows that the above matrices change the coordinates of a vector in V from one basis to another.
- We now discuss change of basis for an operator.

Theorem 10.3. *Suppose $T \in \mathcal{L}(V)$. Let $u_1, \dots, u_n$ and $v_1, \dots, v_n$ be bases of V. Let*

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

Then

$$\mathcal{M}(T, (u_1, \dots, u_n)) = A^{-1}\mathcal{M}(T, (v_1, \dots, v_n))A$$

Proof. We have that

$$\begin{aligned}
\mathcal{M}(T, (u_1, \dots, u_n)) &= \mathcal{M}(IT, (u_1, \dots, u_n), (u_1, \dots, u_n)) \\
&= \mathcal{M}(I, (v_1, \dots, v_n), (u_1, \dots, u_n))\mathcal{M}(T, (u_1, \dots, u_n), (v_1, \dots, v_n)) && \text{Theorem 10.1} \\
&= A^{-1}\mathcal{M}(T, (u_1, \dots, u_n), (v_1, \dots, v_n)) && \text{Theorem 10.2}
\end{aligned}$$

We also have that

$$\begin{aligned}
\mathcal{M}(T, (u_1, \dots, u_n), (v_1, \dots, v_n)) &= \mathcal{M}(TI, (u_1, \dots, u_n), (v_1, \dots, u_n)) \\
&= \mathcal{M}(T, (v_1, \dots, v_n), (v_1, \dots, v_n))\mathcal{M}(I, (u_1, \dots, u_n), (v_1, \dots, v_n)) \\
& && \text{Theorem 10.1} \\
&= \mathcal{M}(T, (v_1, \dots, v_n))A
\end{aligned}$$

Substituting the second equation into the first gives the desired results. ■

- **Trace** (of $T \in \mathcal{L}(V)$, V complex): The sum of the eigenvalues of T with each eigenvalue repeated according to its multiplicity. *Denoted by* $\textbf{trace}\,\boldsymbol{T}$.
- **Trace** (of $T \in \mathcal{L}(V)$, V real): The sum of the eigenvalues of $T_{\mathbb{C}}$ with each eigenvalue repeated according to its multiplicity. *Denoted by* $\textbf{trace}\,\boldsymbol{T}$.
- Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then $\operatorname{trace} T$ equals the negative of the coefficient of z^{n-1} in the characteristic polynomial of T.
- **Trace** (of A): The sum of the diagonal entries of a square matrix A. *Denoted by* $\textbf{trace}\,\boldsymbol{A}$.
- We now build up to proving that $\operatorname{trace} T = \operatorname{trace} \mathcal{M}(T, (v_1, \dots, v_n))$ where $v_1, \dots, v_n$ is an arbitrary basis of V.

Theorem 10.4. *If A and B are square matrices of the same size, then*

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

Proof. Let

$$A = \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{n,1} & \cdots & A_{n,n} \end{pmatrix} \qquad B = \begin{pmatrix} B_{1,1} & \cdots & B_{1,n} \\ \vdots & & \vdots \\ B_{n,1} & \cdots & B_{n,n} \end{pmatrix}$$

The j^{th} diagonal entry of AB is by the definition of matrix multiplication $\sum_{k=1}^{n} A_{j,k}B_{k,j}$. Thus,

$$\begin{aligned}
\operatorname{trace}(AB) &= \sum_{j=1}^{n}\sum_{k=1}^{n} A_{j,k}B_{k,j} \\
&= \sum_{k=1}^{n}\sum_{j=1}^{n} B_{k,j}A_{j,k} \\
&= \operatorname{trace}(BA)
\end{aligned}$$

as desired. ■

- We now prove that the trace of a matrix is unique up to change of basis.

  Theorem 10.5. *Let $T \in \mathcal{L}(V)$. Suppose $u_1, \dots, u_n$ and $v_1, \dots, v_n$ are bases of V. Then*

  $$\operatorname{trace} \mathcal{M}(T, (u_1, \dots, u_n)) = \operatorname{trace} \mathcal{M}(T, (v_1, \dots, v_n))$$

  Proof. Let $A = \mathcal{M}(I, (u_1, \dots, u_n), (v_1, \dots, v_n))$. Then

  $$\begin{aligned} \mathcal{M}(T, (u_1, \dots, u_n)) &= \operatorname{trace}(A^{-1}(\mathcal{M}(T, (v_1, \dots, v_n)))A) && \text{Theorem 10.3} \\ &= \operatorname{trace}((\mathcal{M}(T, (v_1, \dots, v_n)))A^{-1}A) && \text{Theorem 10.4} \\ &= \operatorname{trace} \mathcal{M}(T, (v_1, \dots, v_n)) \end{aligned}$$

  as desired. ■

- We can now prove the main result.

  Theorem 10.6. *Suppose $T \in \mathcal{L}(V)$. Then $\operatorname{trace} T = \operatorname{trace} \mathcal{M}(T)$.*

  Proof. By Theorem 10.5, $\operatorname{trace} \mathcal{M}(T)$ is independent of which basis of V we choose. Thus, to prove that $\operatorname{trace} T = \operatorname{trace} \mathcal{M}(T)$, it will suffice to prove the equality for any basis of V.

  Let $v_1, \dots, v_n$ be the basis of V specified by Theorem 8.13. It follows that $\operatorname{trace} \mathcal{M}(T) = d_1\lambda_1 + \cdots + d_m\lambda_m$ where $\lambda_1, \dots, \lambda_m$ are the eigenvalues of T and $d_1, \dots, d_m$ are there respective multiplicities. But this is just $\operatorname{trace} T$ if V is complex and $\operatorname{trace} T_{\mathbb{C}}$ if V is real, as desired. ■

- Note that the statement of Theorem 10.6 does not specify a basis because trace is invariant under change of basis, as proven in Theorem 10.5.

- The trace is additive.

  Theorem 10.7. *Suppose $S, T \in \mathcal{L}(V)$. Then $\operatorname{trace}(S + T) = \operatorname{trace} S + \operatorname{trace} T$.*

  Proof. We have that

  $$\begin{aligned} \operatorname{trace}(S + T) &= \operatorname{trace} \mathcal{M}(S + T) && \text{Theorem 10.6} \\ &= \operatorname{trace}(\mathcal{M}(S) + \mathcal{M}(T)) \\ &= \operatorname{trace} \mathcal{M}(S) + \operatorname{trace} \mathcal{M}(T) \\ &= \operatorname{trace} S + \operatorname{trace} T && \text{Theorem 10.6} \end{aligned}$$

  as desired. ■

- We now state a curious consequence of the previous theorems that has important applications to quantum theory.

  Theorem 10.8. *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

  $$\begin{aligned} \operatorname{trace}(ST - TS) &= \operatorname{trace}(ST) - \operatorname{trace}(TS) && \text{Theorem 10.7} \\ &= \operatorname{trace} \mathcal{M}(ST) - \operatorname{trace} \mathcal{M}(TS) && \text{Theorem 10.6} \\ &= \operatorname{trace} \mathcal{M}(S)\mathcal{M}(T) - \operatorname{trace} \mathcal{M}(T)\mathcal{M}(S) && \text{Theorem 3.11} \\ &= \operatorname{trace} \mathcal{M}(S)\mathcal{M}(T) - \operatorname{trace} \mathcal{M}(S)\mathcal{M}(T) && \text{Theorem 10.4} \\ &= 0 \end{aligned}$$

  Since $\operatorname{trace} I > 0$ necessarily, $\operatorname{trace}(ST - TS) \neq \operatorname{trace} I$. It follows that $ST - TS \neq I$, as desired. ■

10.B Determinant

- **Determinant** (of $T \in \mathcal{L}(V)$, V complex): The product of the eigenvalues of T with each eigenvalue repeated according to its multiplicity. *Denoted by* $\det T$.
- **Determinant** (of $T \in \mathcal{L}(V)$, V real): The product of the eigenvalues of $T_{\mathbb{C}}$ with each eigenvalue repeated according to its multiplicity. *Denoted by* $\det T$.
- If $\lambda_1, \dots, \lambda_m$ are the distinct eigenvalues of T (or $T_{\mathbb{C}}$ if V is real) with corresponding multiplicities $d_1, \dots, d_m$, then
$$\det T = \lambda_1^{d_1} \cdots \lambda_m^{d_m}$$
- Suppose $T \in \mathcal{L}(V)$. Let $n = \dim V$. Then $\det T$ is $(-1)^n$ times the constant term of the characteristic polynomial of T.
- Invertibility and determinant.

  Theorem 10.9. *An operator on V is invertible if and only if its determinant is nonzero.*

  Proof. Let $T \in \mathcal{L}(V)$. We divide into two cases (V is complex and V is real). Let's begin.

  Suppose first that V is complex. By Theorem 5.7, there is a basis of V with respect to which $\mathcal{M}(T)$ is upper triangular. By Theorem 5.8, T is invertible iff all diagonal entries of $\mathcal{M}(T)$ are nonzero. By Theorem 5.9, all diagonal entries of $\mathcal{M}(T)$ are nonzero iff all eigenvalues of T are nonzero. But this is true iff the product of the eigenvalues of T, i.e., $\det T$ is nonzero, as desired.

  Now suppose that V is real. As before, T is invertible iff 0 is not an eigenvalue of $T_{\mathbb{C}}$. But by Theorem 9.4, it follows in both directions that 0 is not an eigenvalue of T, so $\det T \neq 0$ in this case too, as desired. ∎

- Characteristic polynomial and determinant.

  Theorem 10.10. *Suppose $T \in \mathcal{L}(V)$. Then the characteristic polynomial of T equals* $\det(zI - T)$.

  Proof. Suppose first that V is complex. We know that λ is an eigenvalue of T iff $z - \lambda$ is an eigenvalue of $zI - T$:
$$-(T - \lambda I) = 0 = (zI - T) - (z - \lambda)I$$
  Raising both sides to the $\dim V$ power and taking null spaces proves that the multiplicity of λ wrt. T equals the multiplicity of $z - \lambda$ wrt. $zI - T$. It follows that
$$\det(zI - T) = (z - \lambda_1)^{d_1} \cdots (z - \lambda_n)^{d_n}$$
  which is the characteristic polynomial, as desired.

  The real case follows from applying the complex case to $T_{\mathbb{C}}$. ∎

- **Permutation** (of $(1, \dots, n)$): A list $(m_1, \dots, m_n)$ that contains each of the numbers $1, \dots, n$ exactly once.
- **perm n**: The set of all permutations of $(1, \dots, n)$.
- **Sign** (of a permutation $(m_1, \dots, m_n)$): The number 1 if the number of pairs of integers (j, k) with $1 \leq j < k \leq n$ such that j appears after k in the permutation is even, and the number -1 otherwise (e.g., if the number of such pairs is odd). *Denoted by* **sign n**. *Also known as* **signum**.
  - "In other words, the sign of a permutaton equals 1 if the natural order has been changed an even number of times and equals -1 if the natural order has been changed an odd number of times" (Axler, 2015, p. 313).

- We now prove a connection between the sign and transpositions.

  Theorem 10.11. *Interchanging two entries in a permutation multiplies the sign of the permutation by -1.*

- **Determinant** (of A): The following quantity. *Denoted by* $\det \boldsymbol{A}$. *Given by*

$$\det A = \sum_{(m_1,\dots,m_n)\in \operatorname{perm} n} (\operatorname{sign}(m_1,\dots,m_n))A_{m_1,1}\cdots A_{m_n,n}$$

- We now build up to proving that the determinant of A is invariant with respect to basis.
- Interchanging two columns.

  Theorem 10.12. *Suppose A is a square matrix and B is the matrix obtained from A by interchanging two columns. Then*

$$\det A = -\det B$$

  Proof. Notice that the same products appear in the sum defining the determinants of both matrices. However, the terms appear in different orders; in fact, each term has a unique transposition. Thus, every term of $\det B$ is -1 times the corresponding term in $\det A$ by Theorem 10.11. It follows by factoring out the -1's that $\det A = -\det B$. ■

- If $\mathcal{M}(T)$ has two equal columns, then T is not injective hence not invertible, so $\det = 0$. Similarly...

  Theorem 10.13. *If A is a square matrix that has two equal columns, then* $\det A = 0$.

  Proof. By the definition of A, interchanging the two equal columns of A gives A. But by Theorem 10.12, this implies that

$$\begin{aligned}\det A &= -\det A\\ 2\det A &= 0\\ \det A &= 0\end{aligned}$$

  as desired. ■

- We now generalize Theorem 10.12.

  Theorem 10.14. *Suppose $A = \begin{pmatrix} A_{\cdot,1} & \cdots & A_{\cdot,n}\end{pmatrix}$ is an $n\times n$ matrix and $(m_1,\dots,m_n)$ is a permutation. Then*

$$\det \begin{pmatrix} A_{\cdot,m_1} & \cdots & A_{\cdot,m_n}\end{pmatrix} = (\operatorname{sign}(m_1,\dots,m_n))\det A$$

  Proof. Change A into $\begin{pmatrix} A_{\cdot,m_1} & \cdots & A_{\cdot,m_n}\end{pmatrix}$ iteratively, one column switch at a time, and apply Theorems 10.12 and 10.11. ■

- The determinant is linear.

  Theorem 10.15. *Suppose k,n are positive integers with $1\leq k\leq n$. Fix $n\times 1$ matrices $A_{\cdot,1},\dots,A_{\cdot,n}$ except $A_{\cdot,k}$. Then the function that takes an $n\times 1$ column vector $A_{\cdot,k}$ to*

$$\det \begin{pmatrix} A_{\cdot,1} & \cdots & A_{\cdot,k} & \cdots & A_{\cdot,n}\end{pmatrix}$$

  is a linear map from the vector space of $n\times 1$ matrices with entries in $\mathbb{F}$ to $\mathbb{F}$.

  Proof. The linearity follows from the definition, where each term in the sum contains precisely one entry from the k^{th} column of A. ■

- The determinant of the product of two matrices is equal to the product of the determinants[1].

  Theorem 10.16. *Suppose A, B are square matrices of the same size. Then*

  $$\det(AB) = \det(BA) = (\det A)(\det B)$$

  Proof. Given, but complicated. See LinAlgGIEPNotes on Browne. ∎

- We can now prove that the determinant is independent of basis.

  Theorem 10.17. *Suppose $T \in \mathcal{L}(V)$. Suppose $u_1, \dots, u_n$ and $v_1, \dots, v_n$ are bases of V. Then*

  $$\det \mathcal{M}(T, (u_1, \dots, u_n)) = \det \mathcal{M}(T, (v_1, \dots, v_n))$$

  Proof. Invoke Theorem 10.16 in a paradigm symmetric to that used in the proof of Theorem 10.5. ∎

- We can now prove that the determinant of an operator is equal to the determinant of any of its matrices.

  Theorem 10.18. *Suppose $T \in \mathcal{L}(V)$. Then $\det T = \det \mathcal{M}(T)$.*

  Proof. Invoke Theorem 10.16 in a paradigm symmetric to that used in the proof of Theorem 10.6. ∎

- Like the trace is additive, the determinant is multiplicative.

  Theorem 10.19. *Suppose $S, T \in \mathcal{L}(V)$. Then*

  $$\det(ST) = \det(TS) = (\det S)(\det T)$$

  Proof. Invoke Theorems 10.18 and 10.16 in a paradigm symmetric to that used in the proof of Theorem 10.7. ∎

- We now transition from discussing properties of the determinant to applications.
- Determinant of an isometry.

  Theorem 10.20. *Suppose V is an inner product space and $S \in \mathcal{L}(V)$ is an isometry. Then*

  $$|\det S| = 1$$

  Proof. Suppose first that V is complex. Then by Theorem 7.20, every eigenvalue of S has absolute value 1. Therefore, by the definition of the determinant as the product of the eigenvalues, we have that

  $$|\det S| = |\lambda_1| \cdots |\lambda_m| = 1$$

  as desired.

  Now suppose that V is real. Applying the complexification, we have that $|\det S_{\mathbb{C}}| = 1$ and $\det S = \det S_{\mathbb{C}}$, as desired. ∎

- We have that $\det \sqrt{T^*T} \geq 0$ as a positive operator with all positive eigenvalues.
- We now further investigate the relation between T and $\sqrt{T^*T}$ with respect to the determinant.

  Theorem 10.21. *Suppose V is an inner product space and $T \in \mathcal{L}(V)$. Then*

  $$|\det T| = \det \sqrt{T^*T}$$

[1] The first proof of this theorem was given in 1812 by French mathematicians Jacques Binet and Augustin-Louis Cauchy.

Proof. We have by the Polar Decomposition that there exists an isometry $S \in \mathcal{L}(V)$ such that $T = S\sqrt{T^*T}$. Thus

$$\begin{aligned} |\det T| &= |\det S| \cdot \det \sqrt{T^*T} && \text{Theorem 10.16} \\ &= \det \sqrt{T^*T} && \text{Theorem 10.20} \end{aligned}$$

as desired. ■

- Axler (2015) now discusses applications of the determinant to volume in $\mathbb{R}^n$.
- If $\Omega \subset \mathbb{R}^n$, then the volume of $T(\Omega)$ (where T is a positive operator) equals $\det T$ times the volume of Ω.
- Isometries don't change volume.
- If T is an *arbitrary* operator, then the volume of $T(\Omega)$ equals $|\det T|$ times the volume of Ω.
- Integrals and derivatives are discussed.
- Talks about the Jacobian and change of coordinates.

References

Abbott, E. A. (1952). *Flatland: A romance of many dimensions* (sixth). Dover.
Axler, S. (2015). *Linear algebra done right* (Third). Springer.