LINEAR ALGEBRA, 115A, SPRING 2015

STEVEN HEILMAN

Abstract. These notes are mostly copied from those of T. Tao from 2002, available here.

Contents

1. Introduction, Fields, Vector Spaces, Bases 2
1.1. Introductory Remarks 2
1.3. Fields and Vector Spaces 2
1.4. Three Fundamental Motivations for Linear Algebra 4
1.5. Subspaces, Linear independence 5
1.6. Bases, Spanning Sets 7
1.7. Subspaces and Dimension 13
2. Linear Transformations and Matrices 15
2.2. Linear Transformations 15
2.3. Null spaces, range, coordinate bases 16
2.4. Linear Transformations and Bases 18
2.5. Matrix Representation, Matrix Multiplication 20
2.6. Invertibility, Isomorphism 26
2.7. Change of Coordinates 29
3. Row Operations, The Determinant 30
3.2. Row Operations 30
3.3. Rank of a Matrix 32
3.4. The Determinant 36
4. Eigenvalues, Eigenvectors, Diagonalization 39
4.2. Diagonal Matrices 39
4.3. Eigenvectors and Eigenvalues 39
4.4. Characteristic Polynomial 41
4.5. Diagonalizability 43
5. Inner Products, Adjoints, Spectral Theorems, Self-Adjoint Operators 45
5.2. Inner Product Spaces 45
5.3. Orthogonality 48
5.4. Gram-Schmidt Orthogonalization 50
5.5. Adjoints 55
5.6. Normal Operators 58
5.7. Self-Adjoint Operators 61
5.8. Orthogonal and Unitary Operators (Bonus Section) 63
6. Appendix: Notation 65

Date: February 17, 2017.

1. Introduction, Fields, Vector Spaces, Bases

1.1. **Introductory Remarks.**

1.1.1. *What will we be learning?* We will be learning linear algebra from an abstract perspective.

1.1.2. *Why so abstract?* The abstract approach to learning rigorous mathematics, can be a bit of a difficult adjustment. This approach uses an axiomatic presentation with complete proofs, as opposed to intuitive reasoning and sketches of proofs which are used in your lower division classes. You have probably seen rigorous proofs and the axiomatic method in a class in Euclidean geometry; we will be using this approach for linear algebra. The main proponents of this approach were the Bourbaki group of mainly French mathematicians, starting in the 1930s. The power of the abstract approach is that we can make statements about many examples, simultaneously. The difficulty of the abstract approach is that abstract thinking can require some adjustment for the learner. It is sometimes beneficial to keep some examples in mind to stay grounded, but sometimes these examples can be misleading.

1.2. *A Brief History of Linear Algebra.* Early antecedents for solving systems of linear equations go back at least to Leibniz and Newton. This theory along with matrix theory were developed through the 1800s. Essentially everything that we do in this course was known by the year 1900, though the presentation has been streamlined over the years, as we already discussed. By now matrices are ubiquitous in mathematics. And linear algebra serves as the foundation of quantum mechanics, functional analysis, Fourier analysis, probability theory, partial differential equations, computer science, and several other fields. There is a very good reason that this class is required for all math majors.

1.3. **Fields and Vector Spaces.** In this course, we will be using arithmetic of vectors and fields at an abstract level. For the sake of basic intuition, we can think of a field as $\mathbf{R}$ or $\mathbf{C}$, and we can think of a vector space as $\mathbf{R}^2$ or $\mathbf{R}^n$ for any natural number n with $n \geq 1$. However, many of the statements that we will prove in this course will hold for all objects that satisfy the usual properties of arithmetic with which we are familiar. We formalize these properties below as abstract definitions, when we define both fields and vector spaces, which we will focus on throughout the course.

Definition 1.3.1 (**Binary Operation**)**.** Let F be a set. A **binary operation** is a function $F \times F \to F$.

Example 1.3.2. Addition on the real numbers is a binary operation. Two real numbers (x, y) are mapped to the real number $x + y$.

Definition 1.3.3 (**Field**)**.** A **field** is a set $\mathbf{F}$ with two binary operations $+$ and $\cdot$, such that the following properties hold.

(1) $\forall\ \alpha, \beta \in \mathbf{F}$, $\alpha + \beta = \beta + \alpha$ (commutativity of addition)
(2) $\forall\ \alpha, \beta, \gamma \in \mathbf{F}$, $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$ (associativity of addition)
(3) $\forall\ \alpha, \beta \in \mathbf{F}$, $\alpha \cdot \beta = \beta \cdot \alpha$ (commutativity of multiplication)
(4) $\forall\ \alpha, \beta, \gamma \in \mathbf{F}$, $(\alpha \cdot \beta) \cdot \gamma = \alpha \cdot (\beta \cdot \gamma)$ (associativity of multiplication)
(5) $\forall\ \alpha, \beta, \gamma \in \mathbf{F}$, $\alpha \cdot (\beta + \gamma) = \alpha \cdot \beta + \alpha \cdot \gamma$ (distributivity)
(6) $\exists\ 0 \in \mathbf{F}$ such that $\forall\ \alpha \in \mathbf{F}$, $0 + \alpha = \alpha$ (additive identity)

(7) $\forall\ \alpha \in \mathbf{F}$, $\exists\ -\alpha \in \mathbf{F}$ such that $\alpha + (-\alpha) = 0$ (additive inverse)
(8) $\exists\ 1 \in \mathbf{F}$ such that, $\forall\ \alpha \in \mathbf{F}$, $1 \cdot \alpha = \alpha$ (multiplicative identity)
(9) $\forall\ \alpha \in \mathbf{F}$, $\alpha \neq 0$, $\exists\ \alpha^{-1} \in \mathbf{F}$ such that $\alpha \cdot \alpha^{-1} = 1$ (multiplicative inverse)

Remark 1.3.4. Note that the integers satisfy properties (1) through (8), but not property (9). For all $x \in \mathbf{Z}$, $2x \neq 1$. So, the integers are not a field.

Example 1.3.5. The real numbers $\mathbf{R}$ are a field, with respect to the usual addition and multiplication of real numbers.

Example 1.3.6. The rational numbers $\mathbf{Q}$ are a field, with respect to the usual addition and multiplication of rational numbers.

Example 1.3.7. The set $\mathbf{F} = \{0, 1\}$ can be made into a field if we define addition and multiplication via the following addition and multiplication tables.

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

With these definitions of addition and multiplication, $\mathbf{F}$ is referred to as the field of two elements.

Remark 1.3.8. The elements of a field are often called **scalars**.

Definition 1.3.9 (**Vector Space**)**.** A **vector space** V over a field $\mathbf{F}$ is a set V together with two functions $+\colon V \times V \to V$, $\cdot\colon \mathbf{F} \times V \to V$, such that the following properties hold.

(1) $\forall\ u, v \in V$, $u + v = v + u$ (commutativity of addition)
(2) $\forall\ u, v, w \in V$, $u + (v + w) = (u + v) + w$ (associativity of addition)
(3) $\exists\ 0 \in V$ such that $\forall\ u \in V$, $0 + u = u$ (additive identity)
(4) $\forall\ u \in V$, $\exists\ -u \in V$ such that $u + (-u) = 0$ (additive inverse)
(5) $\forall\ u \in V$, $\forall\ \alpha, \beta \in \mathbf{F}$, $\alpha \cdot (\beta \cdot u) = (\alpha\beta) \cdot u$ (associativity of multiplication)
(6) $\forall\ u \in V$, $\forall\ \alpha, \beta \in \mathbf{F}$, $(\alpha + \beta) \cdot u = \alpha \cdot u + \beta \cdot u$ (scalar distributivity)
(7) $\forall\ u, v \in V$, $\forall\ \alpha \in \mathbf{F}$, $\alpha \cdot (u + v) = \alpha \cdot u + \alpha \cdot v$ (vector distributivity)
(8) $\forall\ u \in V$, $1 \in \mathbf{F}$ satisfies $1 \cdot u = u$ (multiplicative identity)

Remark 1.3.10. Strictly speaking, the field element $0 \in \mathbf{F}$ is distinct from the vector $0 \in V$. However, we use the same notation for both objects, since there is usually no confusion that arises. Yet, at the stage of creating definitions, we should be aware of the difference between these two objects.

Example 1.3.11. $\mathbf{R}$ is a vector space over $\mathbf{R}$.

Example 1.3.12. $\mathbf{R}^2$ is a vector space over $\mathbf{R}$. More generally, for any natural number n, $\mathbf{R}^n$ is a vector space over $\mathbf{R}$. More generally, for any field $\mathbf{F}$, and for any $n \in \mathbf{N}$, $\mathbf{F}^n$ is a vector space over $\mathbf{F}$.

Example 1.3.13. Let x be a real variable. The set $P_2(\mathbf{R})$ of all real polynomials in the variable x of degree at most 2 is a vector space over $\mathbf{R}$. More generally, the set $P(\mathbf{R})$ of all real polynomials in the variable x is a vector space over $\mathbf{R}$. More generally, the set $C^\infty(\mathbf{R})$ of all infinitely differentiable functions in the variable x is a vector space over $\mathbf{R}$.

Remark 1.3.14. Eventually, we will stop writing $\alpha \cdot u$, and we will just write αu, where $\alpha \in \mathbf{F}$ and $u \in V$. No confusion should arise from this change.

To get used to doing proofs, lets prove a fact that follows from the properties of a vector space (Definition 1.3.9).

Proposition 1.3.15 (**Vector Cancellation Law**). *Let V be a vector space over a field $\mathbf{F}$. Let $u, v, w \in V$ such that $u + v = u + w$. Then $v = w$.*

Proof. From property (4) in the Definition of a vector space, there exists $-u \in V$ such that $u + (-u) = 0$. So,

$$\begin{aligned}
v &= 0 + v && \text{, by Property (3) in Definition 1.3.9} \\
&= (u + (-u)) + v \\
&= ((-u) + u) + v && \text{, by Property (1) in Definition 1.3.9} \\
&= (-u) + (u + v) && \text{, by Property (2) in Definition 1.3.9} \\
&= (-u) + (u + w) && \text{, by assumption} \\
&= ((-u) + u) + w && \text{, by Property (2) in Definition 1.3.9} \\
&= (u + (-u)) + w && \text{, by Property (1) in Definition 1.3.9} \\
&= 0 + w \\
&= w && \text{, by Property (3) in Definition 1.3.9.}
\end{aligned}$$

□

After a while we won't do algebraic manipulations in this level of detail. The purpose of the above proof is to get used to justifying each step in our proofs. When doing homework problems, make sure to justify each step of your proof. If you cannot justify each step, then you may have a mistake in your proof!

Exercise 1.3.16. Let V be a vector space over a field $\mathbf{F}$. Using the same level of detail as the proof of Proposition 1.3.15, prove the following facts:

- $\forall\ v \in V$, $0 \cdot v = 0$.
- $\forall\ v \in V$, $(-1) \cdot v = -v$.
- $\forall\ \alpha \in F$, and for $0 \in V$, $\alpha \cdot 0 = 0$.
- $\forall\ \alpha \in \mathbf{F}$, $\forall\ v \in V$, $\alpha \cdot (-v) = (-\alpha) \cdot v = -(\alpha \cdot v)$.

1.4. **Three Fundamental Motivations for Linear Algebra.** We will now present three examples that should motivate the study of linear algebra. Consider the set $X := \{f \in C^\infty([0,1]) \colon f(0) = f(1) = 0\}$. For any $f \in X$, define $Tf := -(d^2/dt^2)f(t)$, where $t \in [0,1]$. Note that X is a vector space over $\mathbf{R}$. We will see later that X is infinite dimensional, so to understand it, we cannot just use our intuition about finite dimensional vector spaces such as $\mathbf{R}^2$. Note that T is linear, in the sense that, for any $f, g \in X$ and for any $\alpha, \beta \in \mathbf{R}$, we have $T(\alpha f + \beta g) = \alpha T(f) + \beta T(g)$. Once again, since X is infinite dimensional, we cannot truly think about T as being a matrix, in the same way that we can understand a linear function on a finite dimensional vector space to be a matrix. However, there are some ways in which we can use our finite dimensional intuition even when X is infinite dimensional. For example, for any $k \geq 0$, $k \in \mathbf{Z}$, the functions $\sin(k\pi t)$ satisfy $T[\sin(k\pi t)] = k^2\pi^2 \sin(k\pi t)$. So, the functions $\sin(k\pi t)$ are eigenfunctions of T with eigenvalues $k^2\pi^2$. And understanding

these eigenfunctions and eigenvalues leads us to an understanding of T. More general linear functions such as T are studied in partial differential equations, quantum mechanics, Fourier analysis, computer science, and so on. The theory of eigenfunctions and eigenvectors from linear algebra can in fact be extended to infinite dimensional vector spaces X. This is done in the mathematical subject of functional analysis. So, for now, we will mostly be studying finite dimensional spaces X, but there is still a lot more to be gained from this theory, as von Neumann and others found in the 1930s.

Linear algebra is also used in search technology, e.g. Google's PageRank algorithm. In this setting, it is desirable to design a large matrix A with very few entries. When we iterate A roughly thirty times to get the matrix A^{30}, then the largest entries of A^{30} give the most relevant websites for a search query. The specific choice of A relies on a linear algebraic interpretation of the set of all websites on the internet. In particular, we take x to be a real vector whose length is the number of websites on the internet, and then A is a square matrix whose side lengths are both the number of websites on the internet. Since A has very few entries, the matrix A^{30} can be computed rather quickly. When Google estimates the time it has taken to complete a search query, it is basically estimating the time it takes to iterate a certain matrix A around 30 times.

Lastly, in sampling and data compression (WAV files, cell phones, JPEG, MPEG, youtube videos,etc.), we once again want to *design* linear transformations which compress data as much as possible. In this setting, a vector x is an audio, image or video file, we design some matrix A in a certain way, and the output Ax is a compressed file. The details of the design of A now come from Fourier analysis.

1.5. **Subspaces, Linear independence.** We are now going to make some definitions that will help us break apart vector spaces into sub-objects. Eventually, we will be able to treat certain vector spaces as sums of simpler pieces. And the simpler pieces (subspaces) will be easier to understand.

Definition 1.5.1 (**Subspace**)**.** Let V be a vector space over a field $\mathbf{F}$, and let $W \subseteq V$ with $W \neq \emptyset$. If W is closed under vector addition and scalar multiplication, we say that W is a **subspace** of V. That is, for all $u, v \in W$, we have $u + v \in W$. And for all $u \in W$, for all $\alpha \in \mathbf{F}$, $\alpha u \in W$.

Remark 1.5.2. If V is a vector space over a field $\mathbf{F}$, and if $W \subseteq V$ is a subspace of V, then W is a vector space over $\mathbf{F}$.

Remark 1.5.3. $C^\infty(\mathbf{R})$ is a subspace of the space of all functions from $\mathbf{R}$ to $\mathbf{R}$.

Remark 1.5.4. Every subspace W of a vector space V must satisfy $0 \in W$. (To see this, choose $\alpha = 0$ in the definition of a subspace.) Note that we do not consider the empty set to be a subspace of V.

The book uses a different definition of a subspace, so let's show that our definition agrees with the definition in the book.

Proposition 1.5.5 (**Subspace Equivalence**)**.** *Let V be a vector space over a field $\mathbf{F}$, and let $W \subseteq V$ with $W \neq \emptyset$. Then W is closed under vector addition and scalar multiplication if and only if W is a vector space over $\mathbf{F}$ (with the operations of addition and scalar multiplication defined on V).*

Proof. We begin with the reverse implication. Suppose W is a vector space over $\mathbf{F}$. Then, from the definition of a vector space (Definition 1.3.9), the operations of addition and multiplication must satisfy $+\colon W \times W \to W$ and $\cdot\colon \mathbf{F} \times W \to W$. That is, W is closed under addition and scalar multiplication.

We now prove the forward implication. Suppose W is closed under vector addition and scalar multiplication. We need to show that W satisfies all of the properties in the definition of a vector space (Definition 1.3.9). Let $u, v, w \in W$, $\alpha, \beta \in \mathbf{F}$. Since $W \subseteq V$, $u, v, w \in V$. Since V is a vector space and $u, v, w \in V$, properties (1), (2), (5), (6), (7) and (8) all apply to u, v, w, α, β. That is, all properties except for properties (3) and (4) must hold for W. So, we will conclude once we show that W satisfies properties (3) and (4). (Note that it is not immediately obvious that $0 \in W$ or $-u \in W$.)

We now show that W satisfies properties (3) and (4). Let $u \in W$. Since $W \subseteq V$, $u \in V$. From Exercise 1.3.16 applied to V, $0 \cdot u = 0$ and $(-1) \cdot u = -u$. Since W is closed under scalar multiplication, we conclude that $0 \in W$ and $-u \in W$. From properties (3) and (4) of Definition 1.3.9 applied to V (recalling that V is a vector space and $u \in V$), we know that $0 + u = u$ and $u + (-u) = 0$. Combining these facts with $0 \in W$ and $-u \in W$, we know that properties (3) and (4) hold for W, as desired. □

Exercise 1.5.6. Show that the intersection of two subspace is also a subspace.

Definition 1.5.7 (**Linear combination**)**.** Let V be a vector space over a field $\mathbf{F}$. Let $u_1, \ldots, u_n \in V$ and let $\alpha_1, \ldots, \alpha_n \in \mathbf{F}$. Then $\sum_{i=1}^{n} \alpha_i u_i$ is called a **linear combination** of the vector elements $u_1, \ldots, u_n$.

Definition 1.5.8 (**Linear dependence**)**.** Let V be a vector space over a field $\mathbf{F}$. Let S be a subset of V. We say that S is **linearly dependent** if there exists a finite set of vectors $u_1, \ldots, u_n \in S$ and there exist $\alpha_1, \ldots, \alpha_n \in \mathbf{F}$ which are not all zero such that $\sum_{i=1}^{n} \alpha_i u_i = 0$.

Definition 1.5.9 (**Linear independence**)**.** Let V be a vector space over a field $\mathbf{F}$. Let S be a subset of V. We say that S is **linearly independent** if S is not linearly dependent.

Example 1.5.10. The set $S = \{(1, 0), (0, 1)\}$ is linearly independent in $\mathbf{R}^2$. The set $S \cup (1, 1)$ is linearly dependent in $\mathbf{R}^2$, since $(1, 0) + (0, 1) - (1, 1) = 0$.

Definition 1.5.11 (**Span**)**.** Let V be a vector space over a field $\mathbf{F}$. Let $S \subseteq V$ be a finite or infinite set. Then the **span** of S, denoted by $\operatorname{span}(S)$, is the set of all finite linear combinations of vectors in S. That is,

$$\operatorname{span}(S) = \left\{ \sum_{i=1}^{n} \alpha_i u_i \colon n \in \mathbf{N}, \alpha_i \in \mathbf{F}, u_i \in S, \forall\, i \in \{1, \ldots, n\} \right\}.$$

Remark 1.5.12. We define $\operatorname{span}(\emptyset) := \{0\}$.

Theorem 1.5.13 (**Span as a Subspace**)**.** *Let V be a vector space over a field $\mathbf{F}$. Let $S \subseteq V$. Then* $\operatorname{span}(S)$ *is a subspace of V such that $S \subseteq \operatorname{span}(S)$. Also, any subspace of V that contains S must also contain* $\operatorname{span}(S)$.

Proof. We first deal with the case that $S = \emptyset$. In this case, $\operatorname{span}(S) = \{0\}$, which is a subspace of V. Also, any subspace contains $\{0\}$, as shown in Remark 1.5.4. Below, we therefore assume that $S \neq \emptyset$.

We first show that $\text{span}(S)$ is a subspace of V.
Step 1. We first show that $\text{span}(S) \subseteq V$. Let $u \in \text{span}(S)$. By the definition of span (Definition 1.5.11), $\exists\ n \in \mathbf{N}$, $\exists\ \alpha_1, \ldots, \alpha_n \in \mathbf{F}$ and $\exists\ u_1, \ldots, u_n \in S \subseteq V$ such that $u = \sum_{i=1}^{n} \alpha_i u_i$. Since V is closed under scalar multiplication and vector addition, we have $u \in V$. Since $u \in \text{span}(S)$ is arbitrary, we conclude that $\text{span}(S) \subseteq V$.
Step 2. We now show that $\text{span}(S)$ is closed under vector addition. Let $v \in \text{span}(S)$. By the definition of span (Definition 1.5.11), $\exists\ m \in \mathbf{N}$, $\exists\ \beta_1, \ldots, \beta_m \in \mathbf{F}$ and $\exists\ v_1, \ldots, v_m \in S \subseteq V$ such that $v = \sum_{i=1}^{m} \beta_i v_i$. So,

$$u + v = \alpha_1 u_1 + \cdots + \alpha_n u_n + \beta_1 v_1 + \cdots + \beta_m v_m.$$

Since $u_1, \ldots, u_n, v_1, \ldots, v_m \in S$, $u + v$ is a linear combination of elements of S. We conclude that $u + v \in \text{span}(S)$. Since $u, v \in \text{span}(S)$ were arbitrary, we have that $\text{span}(S)$ is closed under vector addition.
Step 3. We now show that $\text{span}(S)$ is closed under scalar multiplication. Let $\gamma \in \mathbf{F}$. Recall that $u = \sum_{i=1}^{n} \alpha_i u_i$. Using properties (7) and (5) from the definition of a vector space (Definition 1.3.9),

$$\gamma \cdot u = \gamma \cdot \left(\sum_{i=1}^{n} \alpha_i u_i \right) = \sum_{i=1}^{n} (\gamma \alpha_i) \cdot u_i.$$

That is, $\gamma \cdot u$ is a linear combination of elements of S. Since $u \in \text{span}(S)$ is artbirary, we conclude that $\text{span}(S)$ is closed under scalar multiplication.

Combining Steps 1, 2 and 3 and applying Definition 1.5.1, we get that $\text{span}(S)$ is a subspace of V.

We now show that $S \subseteq \text{span}(S)$. Let $u \in S$. In the definition of the span (Definition 1.5.11), choose $n = 1$, $\alpha_1 = 1$ to get $1 \cdot u \in \text{span}(S)$. By property (8) of the definition of a vector space (Definition 1.3.9), $u = 1 \cdot u \in \text{span}(S)$. Therefore, $S \subseteq \text{span}(S)$.

We now prove the final claim of the Theorem. Let $W \subseteq V$ be a subspace such that $S \subseteq W$. We want to show that $\text{span}(S) \subseteq W$ as well. So, let $n \in \mathbf{N}$, let $u_1, \ldots, u_n \in S$, and let $\alpha_1, \ldots, \alpha_n \in \mathbf{F}$. Since $S \subseteq W$, $u_1, \ldots, u_n \in W$. Since W is a subspace of V, W is closed under scalar multiplication and under vector addition. So, $\sum_{i=1}^{n} \alpha_i u_i \in W$. Since $n \in \mathbf{N}$, $u_1, \ldots, u_n \in S$, and $\alpha_1, \ldots, \alpha_n \in \mathbf{F}$ were arbitrary, we conclude that $\text{span}(S) \subseteq W$. □

1.6. **Bases, Spanning Sets.**

Definition 1.6.1 (**Spanning Set**)**.** Let V be a vector space over a field $\mathbf{F}$. Let $S \subseteq V$. We say that S **spans** V if $\text{span}(S) = V$. In this case, we call S a **spanning set** for V. We can also say that S **generates** V, and S is a **generating set** for V.

Example 1.6.2. The set $\{(1, 0), (0, 1)\}$ is a spanning set for $\mathbf{R}^2$.

Spanning sets S are nice to have, since a spanning set S is sufficient to describe the vector space V (since $\text{span}(S) = V$). If we instead have a set S of linearly dependent vectors, then there is some redundancy in our description of V. To use an analogy, if we want to make a dictionary to describe a language, we want to just make a single entry for each word. It isn't very sensible to have multiple identical entries in our dictionary. The following Theorem then shows that we can remove redundancy in a linearly dependent set of vectors.

Theorem 1.6.3. *Let V be a vector space over a field $\mathbf{F}$. Let $S \subseteq V$ be finite and linearly dependent. Then there exists $u \in S$ such that*

$$\operatorname{span}(S) = \operatorname{span}(S \setminus \{u\}).$$

Conversely, if S is linearly independent and finite, then any proper subset $S' \subsetneq S$ satisfies

$$\operatorname{span}(S') \subsetneq \operatorname{span}(S).$$

Proof. We begin with the first claim. Let $S \subseteq V$ be linearly dependent. Write $S = \{u_1, \ldots, u_n\}$, with $u_i \in V$ for all $i \in \{1, \ldots, n\}$. Since S is linearly dependent, there exist $\alpha_1, \ldots, \alpha_n \in \mathbf{F}$ such that

$$\sum_{i=1}^{n} \alpha_i u_i = 0. \qquad (*)$$

There also exists $i \in \{1, \ldots, n\}$ such that $\alpha_i \neq 0$. By rearranging the vectors $u_1, \ldots, u_n$, we may assume that $\alpha_1 \neq 0$. Then we can rearrange $(*)$ and solve for u_1 to get

$$u_1 = -\alpha_1^{-1} \sum_{i=2}^{n} \alpha_i u_i = \sum_{i=2}^{n} (-\alpha_1^{-1}\alpha_i) u_i. \qquad (**)$$

Since $(S \setminus \{u_1\}) \subseteq S$, $\operatorname{span}(S \setminus \{u_1\}) \subseteq \operatorname{span}(S)$. So, it remains to show that $\operatorname{span}(S \setminus \{u_1\}) \supseteq \operatorname{span}(S)$. To show this, let $w \in \operatorname{span}(S)$. Then there exist $\beta_1, \ldots, \beta_n \in \mathbf{F}$ such that

$$w = \sum_{j=1}^{n} \beta_j u_j.$$

Substituting $(**)$ into this equation,

$$w = \beta_1 \sum_{i=2}^{n} (-\alpha_1^{-1}\alpha_i) u_i + \sum_{j=2}^{n} \beta_j u_j.$$

That is, $w \in \operatorname{span}(S \setminus \{u_1\})$. In conclusion, $\operatorname{span}(S \setminus \{u_1\}) \supseteq \operatorname{span}(S)$, and so $\operatorname{span}(S \setminus \{u_1\}) = \operatorname{span}(S)$.

We now prove the second claim. Since $S' \subseteq S$, $\operatorname{span}(S') \subseteq \operatorname{span}(S)$. So, it remains to find $w \in \operatorname{span}(S)$ such that $w \notin \operatorname{span}(S')$. Since $S' \subsetneq S$, there exists $w \in S$ such that $w \notin S'$. We will show that $w \notin \operatorname{span}(S')$. To show this, we argue by contradiction. Assume that $w \in \operatorname{span}(S')$. Then, there exist $\alpha_1, \ldots, \alpha_n \in \mathbf{F}$ and there exist $u_1, \ldots, u_n \in S' \subseteq S$ such that

$$w = \sum_{i=1}^{n} \alpha_i u_i.$$

That is,

$$0 = (-1)w + \sum_{i=1}^{n} \alpha_i u_i. \qquad (\ddagger)$$

Since $-1 \neq 0$ and $w \notin S'$, we have achieved an equality $(\ddagger)$ that violates the linear independence of S. (If we had $w \in S'$, then the -1 coefficient in front of w could possibly be cancelled by some α_i term in the sum in $(\ddagger)$. And then all coefficients in $(\ddagger)$ could be zero, so $(\ddagger)$ may not give us any linear dependence among elements of S. So, we are really using here that $w \notin S'$.) Since we have achieved a contradiction, we conclude that in fact $w \notin \operatorname{span}(S')$, as desired. $\square$

Exercise 1.6.4. Show that the assumption that S is finite can be removed from the statement of Theorem 1.6.3.

Remark 1.6.5. If we begin with a finite, linearly dependent set of vectors S, then we can apply Theorem 1.6.3 multiple times to eliminate more and more vectors from S to get a linearly independent set.

Definition 1.6.6 (**Basis**)**.** Let V be a vector space over a field $\mathbf{F}$. Let $S \subseteq V$. We say that S is a basis of V if S is a linearly independent set such that $\text{span}(S) = V$.

Example 1.6.7. The set $\{(1,0),(0,1)\}$ is a basis of $\mathbf{R}^2$.

Example 1.6.8. The set $\{1, x, x^2\}$ is a basis of $P_2(\mathbf{R})$.

Example 1.6.9. The set $\{1, x, x^2, x^3, \ldots\}$ is a basis of $P(\mathbf{R})$.

Remark 1.6.10. Bases are the building blocks of a vector space.

Bases are nice for many reasons. One such reason is that they have the following uniqueness property.

Theorem 1.6.11 (**Existence and Uniqueness of Basis Coefficients**)**.** *Let $\{u_1, \ldots, u_n\}$ be a basis for a vector space V over a field $\mathbf{F}$. Then for any vector $u \in V$, there exist unique scalars $\alpha_1, \ldots, \alpha_n \in \mathbf{F}$ such that*

$$u = \sum_{i=1}^{n} \alpha_i u_i.$$

Proof. Let $u \in V$. Since $\{u_1, \ldots, u_n\}$ spans V, there exist scalars $\alpha_1, \ldots, \alpha_n \in \mathbf{F}$ such that

$$u = \sum_{i=1}^{n} \alpha_i u_i. \qquad (*)$$

It remains to show that these scalars are unique. To prove the uniqueness, let $\beta_1, \ldots, \beta_n \in \mathbf{F}$ such that

$$u = \sum_{i=1}^{n} \beta_i u_i. \qquad (**)$$

Subtracting $(*)$ from $(**)$, we get

$$0 = \sum_{i=1}^{n} (\beta_i - \alpha_i) u_i.$$

Since $\{u_1, \ldots, u_n\}$ are linearly independent, we conclude that $(\alpha_i - \beta_i) = 0$ for all $i \in \{1, \ldots, n\}$. That is, $\alpha_i = \beta_i$ for all $i \in \{1, \ldots, n\}$. That is, the scalars $\alpha_1, \ldots, \alpha_n$ are unique. □

Theorem 1.6.12. *Let V be a vector space over a field $\mathbf{F}$. Let S be a linearly independent subset of V. Let $u \in V$ be a vector that does not lie in S.*

(a) *If $u \in \text{span}(S)$, then $S \cup \{u\}$ is linearly dependent, and $\text{span}(S \cup \{u\}) = \text{span}(S)$.*
(b) *If $u \notin \text{span}(S)$, then $S \cup \{u\}$ is linearly independent, and $\text{span}(S \cup \{u\}) \supsetneq \text{span}(S)$.*

Proof of (a). Let $u \in \text{span}(S)$. Then there exist $u_1, \ldots, u_n \in S$, $\alpha_1, \ldots, \alpha_n \in \mathbf{F}$ such that $u = \sum_{i=1}^{n} \alpha_i u_i$. That is,

$$0 = (-1) \cdot u + \sum_{i=1}^{n} \alpha_i u_i.$$

Since $u_i \in S$ for all $i \in \{1, \ldots, n\}$, we conclude that $S \cup \{u\}$ is a linearly dependent set (since $-1 \neq 0$).

Since $S \subseteq S \cup \{u\}$, we know that $\text{span}(S \cup \{u\}) \supseteq \text{span}(S)$. So, it remains to show that $\text{span}(S \cup \{u\}) \subseteq \text{span}(S)$. To this end, let $v \in \text{span}(S \cup \{u\})$. Then there exist $v_1, \ldots, v_m \in S$, $\beta_0, \ldots, \beta_n \in \mathbf{F}$ such that $v = \beta_0 u + \sum_{i=1}^{m} \beta_i v_i$. Since $u = \sum_{i=1}^{n} \alpha_i u_i$, we conclude that

$$v = \beta_0 (\sum_{i=1}^{n} \alpha_i u_i) + \sum_{i=1}^{m} \beta_i v_i.$$

That is, v is a linear combination of elements in S. So, $v \in \text{span}(S)$. In conclusion, $\text{span}(S \cup \{u\}) \subseteq \text{span}(S)$, so $\text{span}(S \cup \{u\}) = \text{span}(S)$. □

Proof of (b). Let $u_1, \ldots, u_n \in S$, and let $\alpha_0, \ldots, \alpha_n \in \mathbf{F}$. Assume that

$$\alpha_0 u + \sum_{i=1}^{n} \alpha_i u_i = 0. \qquad (*)$$

We need to show that $\alpha_0 = \cdots = \alpha_n = 0$. We split into two cases, depending whether or not α_0 is zero. If $\alpha_0 = 0$, then $(*)$ becomes

$$\sum_{i=1}^{n} \alpha_i u_i = 0.$$

And then $\alpha_1 = \cdots = \alpha_n = 0$, since S is linearly independent. On the other hand, if $\alpha_0 \neq 0$, then $(*)$ says

$$u = -\alpha_0^{-1} \left(\sum_{i=1}^{n} \alpha_i u_i \right) = \sum_{i=1}^{n} (-\alpha_0^{-1} \alpha_i) u_i.$$

That is, $u \in \text{span}(S)$, contradicting our assumption that $u \notin \text{span}(S)$. So, we must have $\alpha_0 = 0$, and therefore (as we showed), $\alpha_0 = \alpha_1 = \cdots = \alpha_n = 0$, as desired. Therefore, $S \cup \{u\}$ is linearly independent.

We now prove the second claim of part (b). Since $S \subseteq S \cup \{u\}$, $\text{span}(S \cup \{u\}) \supseteq \text{span}(S)$. Finally, by assumption, $u \notin \text{span}(S)$, so $\text{span}(S \cup \{u\}) \supsetneq \text{span}(S)$, as desired. □

The following theorem elaborates on the previous theorem.

Theorem 1.6.13 (**The Replacement Theorem**). *Let V be a vector space over a field $\mathbf{F}$. Let $S \subseteq V$ be a finite spanning set (i.e. such that $\text{span}(S) = V$). Assume that S has exactly n elements. Let L be a finite subset of V which is linearly independent. Assume that L has exactly m elements. Then $m \leq n$. Moreover, there exists a subset S' of S containing exactly $n - m$ vectors such that $S' \cup L$ spans V.*

Proof. We use induction on m. The base case is $m = 0$, and in this case it is true that $n \geq 0 = m$. Since $\text{span}(S) = V$, we then define $S' := S$, completing the proof.

We now prove the inductive step. Let $m > 0$. Assume that the theorem is true for $m - 1$. Since L has m elements, we can write $L = \{v_1, \ldots, v_m\}$, where $v_i \in V$ for all $i \in \{1, \ldots, m\}$.

Since L is linearly independent, the set $\{v_2, \ldots, v_m\}$ is also linearly independent, by the definition of linear independence. So, by the inductive hypothesis, we apply the theorem to the set of vectors $\{v_2, \ldots, v_m\}$. Then $m-1 \leq n$, and there exists a subset S'' of S containing exactly $n-m+1$ vectors such that $S'' \cup \{v_2, \ldots, v_m\}$ spans V.

Write $S'' = \{w_1, \ldots, w_{n-m+1}\}$, where $w_i \in V$ for all $i \in \{1, \ldots, n-m+1\}$. We now prove that $n \geq m$. We know $n \geq m-1$, so we need to exclude the case $n = m-1$. Since $S'' \cup \{v_2, \ldots, v_m\} = \{w_1, \ldots, w_{n-m+1}, v_2, \ldots, v_m\}$ spans V and $v_1 \in V$, there exist $\alpha_1, \ldots, \alpha_{n-m+1}, \beta_2, \ldots, \beta_m \in \mathbf{F}$ such that

$$v_1 = \alpha_1 w_1 + \cdots + \alpha_{n-m+1} w_{n-m+1} + \beta_2 v_2 + \cdots + \beta_m v_m. \qquad (*)$$

We now argue by contradiction to show that $n \neq m-1$. So, assume to the contrary that $n = m-1$. Then S'' is empty, and $(*)$ becomes

$$v_1 = \beta_2 v_2 + \cdots + \beta_m v_m. \qquad (**)$$

That is,

$$0 = (-1)v_1 + \beta_2 v_2 + \cdots + \beta_m v_m.$$

But $\{v_1, \ldots, v_m\} = L$ is a linearly independent set, so we get a contradiction, since $-1 \neq 0$. We therefore conclude that $n \neq m-1$. Since $n \geq m-1$ also, we conclude that $n \geq m$.

We will now conclude the proof. Since $n \geq m$, and S'' has $n-m+1$ elements, we know that S'' is nonempty. Recall that the set

$$\{w_1, \ldots, w_{n-m+1}, v_2, \ldots, v_m\}$$

spans V, so by adding one vector, we still span V. That is, the set

$$\{w_1, \ldots, w_{n-m+1}, v_1, \ldots, v_m\}$$

spans V. To conclude the proof, we need to remove one of the w_i from this set, and still retain the spanning property.

In equation $(*)$, at least one element of $\alpha_1, \ldots, \alpha_{n-m+1}$ must be nonzero, otherwise we would get $(**)$ and obtain a contradiction. Since the ordering of the vectors in $(*)$ does not matter, we may assume that $\alpha_1 \neq 0$. So, rewriting $(*)$ and solving for w_1,

$$w_1 = v_1 - \alpha_1^{-1}\alpha_2 w_2 - \cdots - \alpha_1^{-1}\alpha_{n-m+1} w_{n-m+1} - \alpha_1^{-1}\beta_2 v_2 - \cdots - \alpha_1^{-1}\beta_m v_m.$$

That is, w_1 is a linear combination of $\{w_2, \ldots, w_{n-m+1}, v_1, \ldots, v_m\}$.

So, define $S' := \{w_2, \ldots, w_{n-m+1}\}$. Then w_1 is a linear combination of elements of $S' \cup L$. By Theorem 1.6.12(a),

$$\operatorname{span}(S' \cup L) = \operatorname{span}(S' \cup L \cup \{w_1\}).$$

But $S' \cup L \cup \{w_1\} = S'' \cup L$, so

$$\operatorname{span}(S' \cup L) = \operatorname{span}(S'' \cup L).$$

Since $S'' \cup \{v_2, \ldots, v_m\}$ spans V and $S'' \cup L \supseteq S'' \cup \{v_2, \ldots, v_m\}$, we conclude that $S'' \cup L$ spans V, so $\operatorname{span}(S' \cup L) = V$. Finally, S' has exactly $n-m$ elements, as desired. $\square$

The Replacement Theorem will allow us to finally start talking about the dimension of finite vector spaces. We now collect some consequences of the Replacement Theorem, some of which will help us in constructing bases of vector spaces.

Corollary 1.6.14. *Let V be a vector space over a field $\mathbf{F}$. Assume that B is a finite basis of V, and B has exactly d elements. Then*

(a) *Any set $S \subseteq V$ containing less than d elements cannot span V. (That is, any spanning set must contain at least d elements.)*
(b) *Any set $S \subseteq V$ containing more than d elements must be linearly dependent. (That is, any linearly independent set in V must contain at most d elements.)*
(c) ***Any basis of V must contain exactly d elements.***
(d) *Any spanning set of V with exactly d elements is a basis of V.*
(e) *Any set of d linearly independent elements of V is a basis of V.*
(f) *Any set of linearly independent elements of V is contained in a basis of V.*
(g) *Any spanning set of V contains a basis.*

Proof of (a). We argue by contradiction. Suppose S spans V, and S has $d' < d$ elements. Since B is linearly independent, the Replacement Theorem (Theorem 1.6.13) implies that $d' \geq d$. But $d' < d$ by assumption. Since we have arrived at a contradiction, we conclude that S cannot span V. □

Proof of (b). First, assume that S is finite. We argue by contradiction. Suppose S is linearly independent, and S has $d' > d$ elements. Since B spans V, the Replacement Theorem (Theorem 1.6.13) implies that $d \geq d'$. But $d' > d$ by assumption. Since we have arrived at a contradiction, we conclude that S is linearly dependent.

Now, assume that S is infinite. Let S' be any subset of S with $d+1$ elements. From what we just proved, we know that S' is linearly dependent. Since $S' \subseteq S$, we conclude that S is linearly dependent. □

Proof of (c). Let $S \subseteq V$ be any basis. Suppose S has d' elements. Since S spans V, $d' \geq d$ by part (a). Since S is linearly independent, $d' \leq d$ by part (b). Therefore $d' = d$. □

Proof of (d). Let $S \subseteq V$ be a spanning set with d elements. It suffices to show that S is linearly independent. To show this, we argue by contradiction. Assume that S is linearly dependent. From Theorem 1.6.3, there exists $u \in S$ such that $S \setminus \{u\}$ is also a spanning set. But $S \setminus \{u\}$ has $d-1$ elements, contradicting part (a). We therefore conclude that S is linearly independent, as desired. □

Proof of (e). Let $S \subseteq V$ be a set of d linearly independent elements. It suffices to show that S is a spanning set. To show this, we argue by contradiction. Suppose S does not span V. Then there exists $u \in V$ such that $u \notin \text{span}(S)$. By Theorem 1.6.12(b), $S \cup \{u\}$ is linearly independent, and it has $d+1$ elements, contradicting part (b) of the present Theorem. We therefore conclude that S is a spanning set. □

Proof of (f). Let $L \subseteq V$ be a set of exactly d' linearly independent elements. By the Replacement Theorem (Theorem 1.6.13), there exists a subset B' of B with exactly $d - d'$ elements such that $L \cup B'$ spans V. Then $L \cup B'$ has at most $(d - d') + d' = d$ elements. Since $L \cup B'$ spans V, $L \cup B'$ must have exactly d elements, by part (a). It remains to show that $L \cup B'$ is linearly independent. This follows from part (d). □

Proof of (g). Let $S \subseteq V$ be a spanning set of V. From part (e), it suffices to find a subset of S of d linearly independent elements. To find such a subset, we argue by contradiction. Suppose every subset of S with d elements has at most $d' < d$ linearly independent elements. Suppose we have d' linearly independent elements $S' := \{u_1, \ldots, u_{d'}\} \subseteq V$. Let $u \in S$ with $u \notin S'$. Then u must be a linear combination of elements of S'. Otherwise, $S \cup \{u\}$ would

be a linearly independent set of $d' + 1$ elements, by Theorem 1.6.12(b). In conclusion, every element of S is a linear combination of elements of S'. So, $\text{span}(S') = \text{span}(S) = V$. So, S' is a basis of V with d' elements. But this fact contradicts part (c), since $d' < d$, and B is a basis of V with d elements. Since we have reached a contradiction, we conclude that there exists a subset S of V with d linearly independent elements, as desired. □

Definition 1.6.15 (**Dimension**)**.** Let V be a vector space over a field $\mathbf{F}$. We say that V is **finite-dimensional** if there exists $d \in \mathbf{N}$ such that V contains a basis with d elements. By Corollary 1.6.14(c), the number d does not depend on the choice of basis of V. We therefore call d the **dimension** of V, and we write $\dim(V) = d$. If the vector space V is not finite-dimensional, we say that V is **infinite-dimensional**, and we write $\dim(V) = \infty$.

Remark 1.6.16. From Corollary 1.6.14(c), we see that a given finite-dimensional vector space V over a field $\mathbf{F}$ has exactly one $d \in \mathbf{N}$ such that V has dimension d. That is, the notion of the dimension of a vector space V over a field $\mathbf{F}$ is well-defined.

Example 1.6.17. $\mathbf{R}^3$ has dimension 3.

Example 1.6.18. $P_2(\mathbf{R})$ has dimension 3.

Example 1.6.19. The vector space $M_{m\times n}(\mathbf{R})$ of $m \times n$ matrices over $\mathbf{R}$ has dimension mn.

Example 1.6.20. $P(\mathbf{R})$ is infinite dimensional.

Example 1.6.21. The complex numbers $\mathbf{C}$ viewed as a vector space over the field $\mathbf{C}$ have dimension 1.

Example 1.6.22. The complex numbers $\mathbf{C}$ viewed as a vector space over the field $\mathbf{R}$ have dimension 2. So, changing the field can change our notion of dimension.

1.7. **Subspaces and Dimension.**

Theorem 1.7.1. *Let V be a finite-dimensional vector space over a field $\mathbf{F}$. Let W be a subspace of V. Then W is also finite-dimensional, and* $\dim(W) \leq \dim(V)$. *Moreover, if* $\dim(W) = \dim(V)$, *then $W = V$.*

Proof. We will build a basis for W. We begin with the zero vector $\{0\}$. If $W = \{0\}$, we stop building the basis. Otherwise, let $u_1 \in W$ be a nonzero vector. If $W = \text{span}(u_1)$, then the basis for W is $\{u_1\}$. Otherwise, let $u_2 \in W$ such that $u_2 \notin \text{span}(u_1)$. By Theorem 1.6.12(b), $\{u_1, u_2\}$ is a linearly independent set. If $W = \text{span}(u_1, u_2)$, then $\{u_1, u_2\}$ is a basis for W, by the definition of basis. Otherwise, let $u_3 \in W$ such that $u_3 \notin \text{span}(u_1, u_2)$. We continue in this way, building this list of vectors. Since V is finite-dimensional, it has a basis consisting of d elements for some $d \in \mathbf{N}$. So, by Corollary 1.6.14(b), we must stop building our list of vectors after at most d steps. Suppose that when this procedure stops, we have n vectors $\{u_1, \ldots, u_n\}$. Then $W = \text{span}(u_1, \ldots, u_n)$, so W is finite-dimensional, $\dim(W) = n$, and $n \leq d$. In the case $n = d$, then $W = \text{span}(u_1, \ldots, u_n)$, and $\{u_1, \ldots, u_n\}$ is a linearly independent set. So, by Corollary 1.6.14(e), $\{u_1, \ldots, u_n\}$ is a basis for V. So, $\text{span}(u_1, \ldots, u_n) = V$, i.e. $W = V$. □

The following result concerning polynomials may appear entirely unrelated to linear algebra. However, if we look at the problem in the right way, the uniqueness statement becomes a relatively easy consequence of the general theory we have developed above.

Theorem 1.7.2 (**Lagrange Interpolation Formula**). *Let $x_1, \ldots, x_n$ be distinct real numbers, and let $y_1, \ldots, y_n \in \mathbf{R}$. Then, there exists a unique polynomial $f \in P_{n-1}(\mathbf{R})$ such that $f(x_i) = y_i$ for all $i \in \{1, \ldots, n\}$. Moreover, for any $x \in \mathbf{R}$, f can be written as*

$$f(x) = \sum_{j=1}^{n} \frac{\prod_{1 \leq k \leq n:\ k \neq j} (x - x_k)}{\prod_{1 \leq k \leq n:\ k \neq j} (x_j - x_k)} \cdot y_j. \qquad (*)$$

Proof. We first show that f is unique. This uniqueness will come from writing f in a suitable basis of $P_{n-1}(\mathbf{R})$, and then applying Theorem 1.6.11.

For each $i \in \{1, \ldots, n\}$, define

$$f_i(x) := \prod_{1 \leq k \leq n:\ k \neq i} \frac{x - x_k}{x_i - x_k}.$$

Note that f_i is a degree $(n-1)$ polynomial,

$$f_i(x_i) = 1, \quad f_i(x_j) = 0 \ \forall\, j \in \{1, \ldots, n\} \setminus \{i\}. \qquad (**)$$

We claim that the set $\{f_i\}_{i=1}^n$ is a basis of $P_{n-1}(\mathbf{R})$. We know that $B := \{1, x, x^2, \ldots, x^{n-1}\}$ is a basis for $P_{n-1}(\mathbf{R})$ with n elements. So, to show that $\{f_i\}_{i=1}^n$ is a basis of $P_{(n-1)}(\mathbf{R})$, it suffices to show that $\{f_i\}_{i=1}^n$ is a linearly independent set, by Corollary 1.6.14(e).

We show that $\{f_i\}_{i=1}^n$ is a linearly independent set by contradiction. Suppose $\{f_i\}_{i=1}^n$ is not linearly independent. Then, there exist $\alpha_1, \ldots, \alpha_n \in \mathbf{R}$ such that

$$\sum_{i=1}^{n} \alpha_i f_i(x) = 0, \quad \forall\, x \in \mathbf{R}, \qquad (\dagger)$$

and there exists $j \in \{1, \ldots, n\}$ such that $\alpha_j \neq 0$. However, using $x = x_j$ in $(\dagger)$, and then applying $(**)$, we get from $(\dagger)$ that $\alpha_j = 0$, a contradiction. We conclude that $\{f_i\}_{i=1}^n$ is a linearly independent set. So, by Theorem 1.6.11, for any $f \in P_{n-1}(\mathbf{R})$, there exist unique scalars $\beta_1, \ldots, \beta_n \in \mathbf{R}$ such that

$$f = \sum_{j=1}^{n} \beta_j f_j. \qquad (\ddagger)$$

If $f \in P_{n-1}(\mathbf{R})$ satisfies $f(x_j) = y_j$ for all $j \in \{1, \ldots, n\}$, we will show that $\beta_j = y_j$ for all $j \in \{1, \ldots, n\}$ in $(\ddagger)$. Fix $j \in \{1, \ldots, n\}$. Using x_j in $(\ddagger)$ and applying $(**)$, we get $f(x_j) = \beta_j$. If $f(x_j) = y_j$ for all $j \in \{1, \ldots, n\}$, we must therefore have $\beta_j = y_j$ in $(\ddagger)$. That is, we exactly recover formula $(*)$.

$$f = \sum_{j=1}^{n} y_j f_j.$$

Finally, note that f defined by the formula $f = \sum_{j=1}^n y_j f_j$ does satisfy $f \in P_{n-1}(\mathbf{R})$ and $f(x_j) = y_j$ for all $j \in \{1, \ldots, n\}$. □

Remark 1.7.3 (An Application to Cryptography.)**.** The following application of Theorem 1.7.2 is known as **Shamir's Secret Sharing**. Suppose I want to have a secret piece of information shared between n people such that all n people can together verify the secret, but any set of $(n-1)$ of the people cannot verify the secret. The following procedure allows us to share the secret in this way. We label the people as integers $i \in \{1, \ldots, n\}$. Let $x_1, \ldots, x_n$ be distinct, nonzero integers, and let $y_1, \ldots, y_n$ be any integers. Each person $i \in \{1, \ldots, n\}$

keeps a value (x_i, y_i). By Theorem 1.7.2, let $f \in P_{n-1}(\mathbf{R})$ be the unique polynomial such that $f(x_i) = y_i$ for all $i \in \{1, \ldots, n\}$. Then the secret information is $f(0)$. To see that the secret cannot be found by $n-1$ people, suppose we only knew the values $\{(x_i, y_i)\}_{i=1}^{n-1}$. Then there would be infinitely many polynomials f such that $f(x_i) = y_i$ for all $i \in \{1, \ldots, n-1\}$ by Theorem 1.7.2. So, the secret $f(0)$ could not be found by $n-1$ people.

2. Linear Transformations and Matrices

2.2. **Linear Transformations.** The general approach to the foundations of mathematics is to study certain spaces, and then to study functions between these spaces. In this course we follow this paradigm. Up until now, we have been studying properties of vector spaces. Vector spaces have a linear structure, and so it is natural to deal with functions between vector spaces that preserve this linear structure. That is, we will concern ourselves with linear transformations between vector spaces. For finite-dimensional spaces, it will turn out that linear transformations can be represented by the action of a matrix on a vector. However, for infinite-dimensional spaces, this representation doesn't quite hold anymore. (Though, thinking by the way of analogy allows many results for infinite-dimensional linear transformations to nearly follow from the finite-dimensional case.) In any case, we can get a good deal of mileage by simply talking about abstract linear transformations, without addressing matrices at all. We will begin this approach below.

Definition 2.2.1. Let V and W be vector spaces over a field $\mathbf{F}$. We call a function $T\colon V \to W$ a **linear transformation** from V to W if, for all $v, v' \in V$ and for all $\alpha \in \mathbf{F}$,

(a) $T(v + v') = T(v) + T(v')$. (T preserves vector addition.)
(b) $T(\alpha v) = \alpha T(v)$. ($T$ preserves scalar multiplication.)

Exercise 2.2.2. Let $T\colon V \to W$ be a linear transformation. Show that $T(0) = 0$.

Example 2.2.3. Define $T(v) := 0$. Then T is linear. This T is known as the zero transformation.

Example 2.2.4. Define $T\colon V \to V$ by $T(v) := v$. Then T is linear.

Example 2.2.5. Define $T\colon \mathbf{R} \to \mathbf{R}$ by $T(x) := x^2$. Then T is **not** linear.

Example 2.2.6. Let $a, b, c, d \in \mathbf{R}$. Define $T\colon \mathbf{R}^2 \to \mathbf{R}^2$ by

$$T\begin{pmatrix} x \\ y \end{pmatrix} := \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Then T is linear.

Example 2.2.7. Define $T\colon C^\infty(\mathbf{R}) \to C^\infty(\mathbf{R})$ by $T(f) := df/dt$. Then T is linear.

Example 2.2.8. Define $T\colon C^\infty(\mathbf{R}) \to C^\infty(\mathbf{R})$ by $T(f) := \int_0^1 f(t)dt$. Then T is linear.

Remark 2.2.9. The set $\mathcal{L}(V, W)$ of all linear transformations from $V \to W$ is itself a vector space over $\mathbf{F}$. We write $\mathcal{L}(V) := \mathcal{L}(V, V)$. Given linear transformations $S, T\colon V \to W$, we define $S + T$ so that, for all $v \in V$, $(S + T)(v) := S(v) + T(v)$. Also, for any $\alpha \in \mathbf{F}$, we define αT so that, for all $v \in V$, $(\alpha T)(v) := \alpha(T(v))$.

2.3. **Null spaces, range, coordinate bases.**

Definition 2.3.1 (**Null Space**)**.** Let V, W be vector spaces over a field $\mathbf{F}$. Let $T\colon V \to W$ be a linear transformation. The **null space** of T, denoted $N(T)$, is defined as

$$N(T) := \{v \in V\colon T(v) = 0\}.$$

Remark 2.3.2. $N(T)$ is also referred to as the **kernel** of T. Note that $N(T)$ is a subspace of V, so its dimension can be defined.

Definition 2.3.3 (**Nullity**)**.** Let V, W be vector spaces over a field $\mathbf{F}$. Let $T\colon V \to W$ be a linear transformation. The **nullity** of T, denoted nullity(T), is defined as

$$\dim(N(T)).$$

Theorem 2.3.4. *Let V, W be vector spaces over a field $\mathbf{F}$. Let $T\colon V \to W$ be a linear transformation. Then T is injective if and only if $N(T) = \{0\}$.*

Proof. Suppose T is injective. We will show that $N(T) = \{0\}$. Note that $T(0) = 0$ by Exercise 2.2.2, so $\{0\} \subseteq N(T)$. It now suffices to show that $N(T)$ has only one element, which we prove by contradiction. Suppose there exist $v, v' \in N(T)$ such that $v \neq v'$. Since T is injective, $T(v) \neq T(v')$. But $v, v' \in N(T)$ imply $0 = T(v) = T(v')$, a contradiction. We conclude that $N(T)$ has only one element, as desired.

Now, suppose $N(T) = \{0\}$. We will show that T is injective. Let $v, v' \in V$ such that $T(v) = T(v')$. By linearity of T, $T(v - v') = T(v) - T(v') = 0$, so $v - v' \in N(T)$. Since $N(T) = \{0\}$, $v - v' = 0$, so that $v = v'$, proving the injectivity of T. □

Definition 2.3.5 (**Range**)**.** Let $T\colon V \to W$ be a linear transformation. The **range** of T, denoted $R(T)$, is defined as

$$R(T) := \{T(v)\colon v \in V\}.$$

Remark 2.3.6. Note that $R(T)$ is a subspace of W, so its dimension can be defined.

Definition 2.3.7 (**Rank**)**.** Let V, W be vector spaces over a field $\mathbf{F}$. Let $T\colon V \to W$ be a linear transformation. The **rank** of T, denoted rank(T), is defined as

$$\dim(R(T)).$$

Exercise 2.3.8. Let $T\colon V \to W$ be a linear transformation. Prove that $N(T)$ is a subspace of V and that $R(T)$ is a subspace of W.

Theorem 2.3.9 (**Dimension Theorem/ Rank-Nullity Theorem**)**.** *Let V, W be vector spaces over a field $\mathbf{F}$. Let $T\colon V \to W$ be linear. If V is finite-dimensional, then*

$$\text{nullity}(T) + \text{rank}(T) = \dim(V).$$

Proof. Since V is finite dimensional, and $N(T) \subseteq V$ is a subspace, $N(T)$ is finite dimensional by Theorem 1.7.1. In particular, a basis $\{v_1, \ldots, v_k\}$ for $N(T)$ exists, by the definition of finite-dimensionality. So, the set $\{v_1, \ldots, v_k\} \subseteq V$ is linearly independent. By Corollary 1.6.14(f), the set $\{v_1, \ldots, v_k\}$ is therefore contained in a basis for V. (Since V is finite-dimensional, a basis for V exists, so we can apply Corollary 1.6.14.) So, we have a basis $\{v_1, \ldots, v_k, u_1, \ldots, u_m\}$ for V. That is, nullity$(T) = k$ and $\dim(V) = k + m$. It remains to show that rank$(T) = m$.

We now show that $\text{rank}(T) = m$. To show this, it suffices to show that $\{Tu_1, \ldots, Tu_m\}$ is a basis for $R(T)$. Let us therefore show that $\{Tu_1, \ldots, Tu_m\}$ is a linearly independent set. We prove this by contradiction. Suppose $\{Tu_1, \ldots, Tu_m\}$ is not a linearly independent set. Then there exist $\alpha_1, \ldots, \alpha_m \in \mathbf{F}$ which are not all equal to zero, such that

$$\sum_{i=1}^{m} \alpha_i T u_i = 0.$$

Since T is linear, we can rewrite this as

$$T\left(\sum_{i=1}^{m} \alpha_i u_i\right) = 0.$$

That is, $\sum_{i=1}^{m} \alpha_i u_i \in N(T)$. Since $\{v_1, \ldots, v_k\}$ is a basis for $N(T)$, there exist scalars $\beta_1, \ldots, \beta_k \in \mathbf{F}$ such that

$$\sum_{i=1}^{m} \alpha_i u_i = \sum_{i=1}^{k} \beta_i v_i.$$

That is,

$$\sum_{i=1}^{m} \alpha_i u_i - \sum_{i=1}^{k} \beta_i v_i = 0. \qquad (*)$$

Since the set $\{v_1, \ldots, v_k, u_1, \ldots, u_m\}$ is a basis for V, this set is linearly independent. So all the coefficients in $(*)$ are zero. In particular, $\alpha_1 = \cdots = \alpha_m = 0$. But we assumed that some α_i was nonzero. Since we have achieved a contradiction, we conclude that $\{Tu_1, \ldots, Tu_m\}$ is a linearly independent set.

It now remains to show that $\{Tu_1, \ldots, Tu_m\}$ is a spanning set of $R(T)$. Let $w \in R(T)$. We need to show that w is a linear combination of $\{Tu_1, \ldots, Tu_m\}$. Since $w \in R(T)$, there exists $u \in V$ such that $T(u) = w$. Since $\{v_1, \ldots, v_k, u_1, \ldots, u_m\}$ is a basis for V, there exist scalars $\gamma_1, \ldots, \gamma_k, \delta_1, \ldots, \delta_m \in \mathbf{F}$ such that $u = \sum_{i=1}^{k} \gamma_i v_i + \sum_{i=1}^{m} \delta_i u_i$. Applying T to both sides of this equation, and recalling that $v_i \in N(T)$ for all $i \in \{1, \ldots, k\}$, we get

$$T(u) = T\left(\sum_{i=1}^{k} \gamma_i v_i + \sum_{i=1}^{m} \delta_i u_i\right) = T\left(\sum_{i=1}^{m} \delta_i u_i\right) = \sum_{i=1}^{m} \delta_i T(u_i). \qquad (**)$$

Since $w = T(u)$, we have just expressed w as a linear combination of $\{Tu_1, \ldots, Tu_m\}$, as desired. We conclude that $\{Tu_1, \ldots, Tu_m\}$ is a spanning set for $R(T)$, so that $\text{rank}(T) = m$, as desired. □

Lemma 2.3.10. *Let V and W be finite-dimensional vector spaces over a field $\mathbf{F}$. Assume that $\dim(V) = \dim(W)$. Let $T\colon V \to W$ be linear. Then T is one-to-one if and only if T is onto.*

Proof. We only prove the forward implication. Suppose T is one-to-one. Then $N(T) = \{0\}$ by Theorem 2.3.4. By the Dimension Theorem (Theorem 2.3.9), $\text{rank}(T) = \dim(V)$. Since $\dim(V) = \dim(W)$, $\text{rank}(T) = \dim(W)$. Since $R(T)$ is a subspace of W, and $\dim(R(T)) = \dim(W)$, we conclude that $R(T) = W$ by Theorem 1.7.1. So, T is onto, as desired. □

Exercise 2.3.11. Prove the reverse implication of Lemma 2.3.10.

Exercise 2.3.12. Define $T\colon C(\mathbf{R}) \to C(\mathbf{R})$ by $Tf(x) := \int_0^x f(t)dt$. Note that T is linear and one-to-one, but not onto, since there does not exist $f \in C(\mathbf{R})$ such that $T(f)(x) = 1$ for all $x \in \mathbf{R}$. Define $S\colon P(\mathbf{R}) \to P(\mathbf{R})$ by $Sf := df/dt$. Note that S is linear and onto, but S is not one-to-one, since S maps the constant function 1 to the zero function. How can you reconcile these facts with Lemma 2.3.10?

2.4. **Linear Transformations and Bases.** We will now isolate a few facts related to the main steps of the proof of the Dimension Theorem. These facts will be useful for us in our later discussion of isomorphism.

Theorem 2.4.1. *Let V, W be vector spaces over a field $\mathbf{F}$. Let $T\colon V \to W$ be a linear transformation. Assume that $\{v_1, \ldots, v_n\}$ spans V. Then $\{Tv_1, \ldots, Tv_n\}$ spans $R(T)$.*

Proof. Let $w \in R(T)$. We need to express w as a linear combination of $\{Tv_1, \ldots, Tv_n\}$. Since $w \in R(T)$, there exists $v \in V$ such that $T(v) = w$. Since $\{v_1, \ldots, v_n\}$ spans V, there exist scalars $\alpha_1, \ldots, \alpha_n \in \mathbf{F}$ such that $v = \sum_{i=1}^n \alpha_i v_i$. Applying T to both sides of this equality, and then using linearity of T,

$$T(v) = T\left(\sum_{i=1}^n \alpha_i v_i\right) = \sum_{i=1}^n \alpha_i T(v_i).$$

Since $w = T(v)$, we have expressed w as a linear combination of $\{Tv_1, \ldots, Tv_n\}$, as desired. □

Theorem 2.4.2. *Let V, W be vector spaces over a field $\mathbf{F}$. Let $T\colon V \to W$ be a linear transformation which is one-to-one. Assume that $\{v_1, \ldots, v_n\}$ is linearly independent. Then $\{T(v_1), \ldots, T(v_n)\}$ is also linearly independent.*

Proof. We argue by contradiction. Assume that $\{T(v_1), \ldots, T(v_n)\}$ is linearly dependent. Then there exist scalars $\alpha_1, \ldots, \alpha_n \in \mathbf{F}$ not all equal to zero such that $\sum_{i=1}^n \alpha_i T(v_i) = 0$. Applying linearity of T, this equation says $T(\sum_{i=1}^n \alpha_i v_i) = 0$. Since T is one-to-one, we must have

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

However, the set $\{v_1, \ldots, v_n\}$ is linearly independent, so we must have $\alpha_1 = \cdots = \alpha_n = 0$. But at least one α_i must be nonzero, a contradiction. We conclude that $\{T(v_1), \ldots, T(v_n)\}$ is linearly independent, as desired. □

Corollary 2.4.3 (**Bijections Preserve Bases**). *Let V, W be vector spaces over a field $\mathbf{F}$. Let $T\colon V \to W$ be a linear transformation which is one-to-one and onto. Assume that $\{v_1, \ldots, v_n\}$ is a basis for V. Then $\{T(v_1), \ldots, T(v_n)\}$ is a basis for W. And therefore,* $\dim(V) = \dim(W) = n$.

Proof. Since $\{v_1, \ldots, v_n\}$ is a basis for V, $\{v_1, \ldots, v_n\}$ spans V. So, from Theorem 2.4.1, $\{T(v_1), \ldots, T(v_n)\}$ spans $R(T)$. Since T is onto, $R(T) = W$, so $\{T(v_1), \ldots, T(v_n)\}$ spans W. It remains to show that $\{T(v_1), \ldots, T(v_n)\}$ is linearly independent. Since $\{v_1, \ldots, v_n\}$ is a basis for V, $\{v_1, \ldots, v_n\}$ is linearly independent. So, from Theorem 2.4.2, $\{T(v_1), \ldots, T(v_n)\}$ is linearly independent, as desired. □

As we now show, if $T\colon V \to W$ is linear and T is defined only on a basis of V, then this is sufficient to define T over all vectors in V. We phrase this theorem as a combined existence and uniqueness statement.

Theorem 2.4.4 (**Rigidity of Linear Transformations**). *Let V, W be vector spaces over a field $\mathbf{F}$. Assume that $\{v_1, \ldots, v_n\}$ is a basis for V. Let $\{w_1, \ldots, w_n\}$ be any vectors in W. Then there exists a unique linear transformation $T\colon V \to W$ such that $T(v_i) = w_i$ for all $i \in \{1, \ldots, n\}$.*

Proof. We first prove that T exists. Let $v \in V$. From Theorem 1.6.11, there exist unique scalars $\alpha_1, \ldots, \alpha_n \in \mathbf{F}$ such that $v = \sum_{i=1}^n \alpha_i v_i$. Suppose we define a map

$$T(\sum_{i=1}^n \alpha_i v_i) := \sum_{i=1}^n \alpha_i w_i. \qquad (*)$$

Observe that $T\colon V \to W$ is a map. In particular, since the scalars $\alpha_1, \ldots, \alpha_n \in \mathbf{F}$ depend uniquely on v, T is well-defined. We now check that $T(v_i) = w_i$ for all $i \in \{1, \ldots, n\}$. Note that

$$v_i = 1 \cdot v_i + \sum_{1 \leq j \leq n\colon\, j \neq i} 0 \cdot v_j.$$

So, plugging this formula into $(*)$ shows that $T(v_i) = w_i$.

We now need to verify that T is linear. Let $\alpha \in \mathbf{F}$. We first verify that $T(\alpha v) = \alpha T(v)$.

$$\begin{aligned} T(\alpha v) = T\left(\sum_{i=1}^n (\alpha\alpha_i) v_i\right) &= \sum_{i=1}^n \alpha\alpha_i w_i \qquad \text{, by } (*) \\ &= \alpha(\sum_{i=1}^n \alpha_i w_i) = \alpha T(v) \qquad \text{, by } (*). \end{aligned}$$

So, $T(\alpha v) = \alpha T(v)$ for all $v \in V$ and for all $\alpha \in \mathbf{F}$. Let $v' \in V$. We now verify that $T(v + v') = T(v) + T(v')$. There exist unique scalars $\beta_1, \ldots, \beta_n \in \mathbf{F}$ such that $v' = \sum_{i=1}^n \beta_i v_i$. We now check

$$\begin{aligned} T(v + v') = T\left(\sum_{i=1}^n (\alpha_i + \beta_i) v_i\right) &= \sum_{i=1}^n (\alpha_i + \beta_i) w_i \qquad \text{, by } (*) \\ &= \sum_{i=1}^n \alpha_i w_i + \sum_{i=1}^n \beta_i w_i = T(v) + T(v') \qquad \text{, by } (*). \end{aligned}$$

In conclusion, the map T defined by $(*)$ is in fact a linear transformation.

We now finish the proof by showing that T is unique. Suppose some other linear transformation $T'\colon V \to W$ satisfies $T(v_i) = w_i$ for all $i \in \{1, \ldots, n\}$. Then $(T - T')(v_i) = 0$ for all $i \in \{1, \ldots, n\}$. So, for any $v \in V$, once we write $v = \sum_{i=1}^n \alpha_i v_i$, we have by linearity of $(T - T')$

$$(T - T')(v) = (T - T')(\sum_{i=1}^n \alpha_i v_i) = \sum_{i=1}^n \alpha_i (T - T')(v_i) = 0.$$

That is, $T - T' = 0$, so $T = T'$, as desired. $\square$

2.5. **Matrix Representation, Matrix Multiplication.**

Definition 2.5.1 (**Ordered Basis**)**.** Let V be a finite-dimensional vector space over a field $\mathbf{F}$. An **ordered basis** for V is a an ordered set $(v_1, \ldots, v_n)$ of elements of V such that $\{v_1, \ldots, v_n\}$ is a basis of V.

Example 2.5.2. One ordered basis for $\mathbf{R}^2$ is $((1,0),(0,1))$.

Definition 2.5.3 (**Coordinate Vector**)**.** Let $\beta = (v_1, \ldots, v_n)$ be an ordered basis for V, and let $v \in V$. From Theorem 1.6.11, there exist unique scalars such that $v = \sum_{i=1}^{n} \alpha_i v_i$. The scalars $\alpha_1, \ldots, \alpha_n$ are referred to as the **coordinates** of v with respect to β. We then define the **coordinate vector** of v relative to β by

$$[v]^\beta := \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

Example 2.5.4. Let $v := (3,4)$. If $\beta = ((1,0),(0,1))$. Then $v = 3(1,0) + 4(0,1)$, so

$$[v]^\beta = \begin{pmatrix} 3 \\ 4 \end{pmatrix}.$$

If $\beta' = ((1,-1),(1,1))$, then $v = (-1/2)(1,-1) + (7/2)(1,1)$, so

$$[v]^{\beta'} = \begin{pmatrix} -1/2 \\ 7/2 \end{pmatrix}.$$

If $\beta'' = ((3,4),(0,1))$, then $v = 1(3,4) + 0(0,1)$

$$[v]^{\beta''} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Definition 2.5.5 (**Matrix Representation**)**.** Let V, W be finite-dimensional vector spaces. Let $\beta = (v_1, \ldots, v_n)$ be an ordered basis for V, and let $\gamma = (w_1, \ldots, w_m)$ be an ordered basis for W. Let $T \colon V \to W$ be linear. Then, for each $j \in \{1, \ldots, n\}$, there exist unique scalars $a_{1j}, \ldots, a_{mj} \in \mathbf{F}$ by Theorem 1.6.11 such that

$$T(v_j) = \sum_{i=1}^{m} a_{ij} w_i.$$

We therefore define the **matrix representation** of T with respect to the bases β and γ by

$$[T]_\beta^\gamma = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

Remark 2.5.6. Note that the j^{th} column of $[T]_\beta^\gamma$ is exactly $[T(v_j)]^\gamma$, so that

$$[T]_\beta^\gamma = ([T(v_1)]^\gamma, [T(v_2)]^\gamma, \ldots, [T(v_n)]^\gamma).$$

So, if we have an arbitrary $v \in V$, and we write v uniquely as $v = \sum_{j=1}^{n} b_j v_j$ where $b_1, \ldots, b_n \in \mathbf{F}$, then by linearity, $Tv = \sum_{j=1}^{n} b_j T(v_j)$. That is,

$$Tv = \sum_{j=1}^{n} \sum_{i=1}^{m} b_j a_{ij} w_i = \sum_{i=1}^{m} (\sum_{j=1}^{n} b_j a_{ij}) w_i$$

If we also express Tv in the basis γ, so that $Tv = \sum_{i=1}^{m} c_i w_i$ where $c_1, \ldots, c_m \in \mathbf{F}$, then we equate like terms to get $c_i = \sum_{j=1}^{n} b_j a_{ij}$ for all $1 \leq i \leq m$. In matrix form, this becomes

$$\begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

Or, using our ordered basis notation,

$$[Tv]^{\gamma} = [T]^{\gamma}_{\beta} [v]^{\beta}.$$

Remark 2.5.7. The important point here is that a linear transformation $T\colon V \to W$ has a meaning that does not depend on any ordered basis. However, when we view T from different perspectives (i.e. we examine $[T]^{\gamma}_{\beta}$ for different ordered bases β, γ), then T may look very different. One of the major goals of linear algebra is to take a T and view it from the "correct" basis, so that $[T]^{\gamma}_{\beta}$ takes a rather simple form, and therefore T becomes easier to understand. For example, if we could find ordered bases β, γ such that $[T]^{\gamma}_{\beta}$ becomes a diagonal matrix, then this would be really nice, since diagonal matrices are fairly easy to understand, and therefore we would better understand T. Unfortunately, we cannot always find bases such that $[T]^{\gamma}_{\beta}$ becomes diagonal, but in certain cases this can be done.

Remark 2.5.8. If we have a linear transformation $T\colon V \to W$, then specifying ordered bases β, γ gives a matrix representation $[T]^{\gamma}_{\beta}$. Conversely, if we have a matrix representation $[T]^{\gamma}_{\beta}$, then we know how T acts on an ordered basis. So, by Theorem 2.4.4, we can recover $T\colon V \to W$ from the matrix representation $[T]^{\gamma}_{\beta}$.

Example 2.5.9. Let $T\colon \mathbf{R}^2 \to \mathbf{R}^2$ be the linear transformation that takes any vector $(x, y) \in \mathbf{R}^2$ and rotates this vector counterclockwise around the origin by an angle $\pi/2$. Note that this description of T does not make use of any ordered basis. Let us find two different matrix representations of T. We first use $\beta = \gamma = ((1,0),(0,1))$. In this case, note that $T(1,0) = (0,1)$ and $T(0,1) = (-1,0)$. So, $T(1,0) = 0(1,0) + 1(0,1)$ and $T(0,1) = -1(1,0) + 0(0,1)$, and

$$[T]^{\gamma}_{\beta} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

We will now find a matrix representation of T that is the identity matrix. Let $\beta := ((1,0),(0,1))$ and let $\gamma := ((0,1),(-1,0))$. Then $T(1,0) = 1(0,1) + 0(-1,0)$ and $T(0,1) = 0(0,1) + 1(-1,0)$, so

$$[T]^{\gamma}_{\beta} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Recall that, in Remark 2.2.9, we noted that the set $\mathcal{L}(V, W)$ of all linear transformations from $V \to W$ is itself a vector space over $\mathbf{F}$. Given linear transformations $S, T \colon V \to W$, we defined $S + T$ so that, for all $v \in V$, $(S + T)(v) := S(v) + T(v)$. Also, for any $\alpha \in \mathbf{F}$, we defined αT so that, for all $v \in V$, $(\alpha T)(v) := \alpha(T(v))$. We can also define the product, or composition, or linear transformations as follows.

Definition 2.5.10 (**Product/Composition**)**.** Let U, V, W be vector spaces over a field $\mathbf{F}$. Let $S \colon V \to W$ and let $T \colon U \to V$ be linear transformations. We define the **product** or **composition** $ST \colon U \to W$ by the formula

$$ST(u) := S(T(u)) \qquad \forall\, u \in U.$$

Exercise 2.5.11. Using the linearity of S and T, show that $ST \colon U \to W$ is a linear transformation.

Definition 2.5.12 (**Matrix Multiplication**)**.** Let A be an $m \times \ell$ matrix, and let B be an $n \times m$ matrix. That is, A is a collection of scalars arranged into m rows and ℓ columns as follows

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1\ell} \\ A_{21} & A_{22} & \cdots & A_{2\ell} \\ \vdots & \vdots & \cdots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{m\ell} \end{pmatrix}.$$

Then the $n \times \ell$ matrix BA is defined, so that the (k, i) entry of BA is given by

$$(BA)_{ki} := \sum_{j=1}^{m} B_{kj} A_{ji}. \qquad 1 \leq k \leq n,\, 1 \leq i \leq \ell$$

Definition 2.5.13. The $n \times n$ identity matrix I_n is defined by

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

Note that the composition of two linear transformations evidently has a natural definition in Definition 2.5.10. On the other hand, matrix multiplication in Definition 2.5.12 may have appeared somewhat unnatural at first sight. So, perhaps surprisingly, we now show that the composition of linear transformations exactly defines the matrix multiplication to which we are accustomed. Put another way, the matrix multiplication in Definition 2.5.12 is a realization, in coordinates, of the composition of two linear transformations.

Theorem 2.5.14 (**Equivalence of Composition and Matrix Multiplication**)**.** *Suppose U, V, W are vector spaces over a field $\mathbf{F}$. Let $S \colon V \to W$ and let $T \colon U \to V$ be linear transformations. Assume that U is ℓ-dimensional and it has an ordered basis $\alpha = (u_1, \ldots, u_\ell)$. Assume that V is m-dimensional and it has an ordered basis $\beta = (v_1, \ldots, v_m)$. Assume that W is n-dimensional and it has an ordered basis $\gamma = (w_1, \ldots, w_n)$. Then*

$$[ST]_\alpha^\gamma = [S]_\beta^\gamma [T]_\alpha^\beta.$$

Proof. We first apply Definition 2.5.5 to T. Then there exist scalars $\{a_{ji}\}_{1\leq j\leq m, 1\leq i\leq \ell}$ such that, for each $1 \leq i \leq \ell$,

$$T(u_i) = \sum_{j=1}^{m} a_{ji} v_j. \tag{2.5.1}$$

That is,

$$[T]_{\alpha}^{\beta} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1\ell} \\ a_{21} & a_{22} & \cdots & a_{2\ell} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{m\ell} \end{pmatrix}. \tag{2.5.2}$$

We now apply Definition 2.5.5 to S. Then there exist scalars $\{b_{kj}\}_{1\leq k\leq n, 1\leq j\leq m}$ such that, for each $1 \leq j \leq m$,

$$S(v_j) = \sum_{k=1}^{n} b_{kj} w_k. \tag{2.5.3}$$

That is,

$$[S]_{\beta}^{\gamma} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nm} \end{pmatrix}. \tag{2.5.4}$$

Applying S to both sides of (2.5.1) and using linearity of S,

$$\begin{aligned} S(T(u_i)) = S\left(\sum_{j=1}^{m} a_{ji} v_j\right) &= \sum_{j=1}^{m} a_{ji} S(v_j) \\ &= \sum_{j=1}^{m} a_{ji} \sum_{k=1}^{n} b_{kj} w_k \qquad \text{, by (2.5.3).} \end{aligned}$$

Changing the order of summation, we get

$$ST(u_i) = \sum_{k=1}^{n} \left(\sum_{j=1}^{m} b_{kj} a_{ji}\right) w_k. \tag{2.5.5}$$

So, for each $1 \leq k \leq n$ and $1 \leq i \leq \ell$, define

$$c_{ki} := \sum_{j=1}^{m} b_{kj} a_{ji}. \tag{2.5.6}$$

Then (2.5.5) becomes

$$ST(u_i) = \sum_{k=1}^{n} c_{ki} w_k. \tag{2.5.7}$$

That is, using the definitions of α and γ,

$$[ST]_\alpha^\gamma = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1\ell} \\ c_{21} & c_{22} & \cdots & c_{2\ell} \\ \vdots & \vdots & \cdots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{n\ell} \end{pmatrix}. \tag{2.5.8}$$

Finally, we use (2.5.2) and (2.5.4), and then perform the matrix multiplication

$$[S]_\beta^\gamma [T]_\alpha^\beta = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nm} \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1\ell} \\ a_{21} & a_{22} & \cdots & a_{2\ell} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{m\ell} \end{pmatrix}. \tag{2.5.9}$$

Then the matrix multiplication in (2.5.9), defined in Definition 2.5.12, agrees with the matrix in (2.5.8), because of (2.5.6). That is, $[ST]_\alpha^\gamma = [S]_\beta^\gamma [T]_\alpha^\beta$, as desired. □

2.5.1. *Matrices as Linear Transformations.* We showed in Theorem 2.5.14 that composing two linear transformations is equivalent to using matrix multiplication. We now belabor this point by beginning with a matrix, and then using the theory of linear transformations to prove associativity of matrix multiplication. We could prove that matrix multiplication is associative by taking three matrices and then writing out all the relevant terms. However, the "coordinate-free" approach below ends up being a bit more elegant. This proof strategy is part of a larger paradigm, in which "coordinate-free" proofs end up being more enlightening that coordinate-reliant proofs.

Definition 2.5.15. Consider the vector space $\mathbf{F}^n$ over the field $\mathbf{F}$. The **standard basis** for $\mathbf{F}^n$ is defined as

$$(e_1, \ldots, e_n) = ((1, 0, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)).$$

Definition 2.5.16. Let A be an $m \times n$ matrix of scalars in a field $\mathbf{F}$. Define $L_A \colon \mathbf{F}^n \to \mathbf{F}^m$ by the formula

$$L_A(u) := Au, \quad \forall\, u \in \mathbf{F}^n.$$

Here we think of vectors in $\mathbf{F}^n$ and $\mathbf{F}^m$ as column vectors. Note that L_A is linear.

Lemma 2.5.17. *Let α be the standard basis of $\mathbf{F}^n$ and let β be the standard basis of $\mathbf{F}^m$. Let $A \in M_{m\times n}(\mathbf{F})$. Then $[L_A]_\alpha^\beta = A$. Let $T \colon \mathbf{F}^n \to \mathbf{F}^m$ be a linear transformation. Then $L_{[T]_\alpha^\beta} = T$.*

Proof. Let $u \in \mathbf{F}^n$ be a column vector. That is, there exist $\alpha_1, \ldots, \alpha_n \in \mathbf{F}^n$ such that

$$u = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

That is, $u = \sum_{i=1}^n \alpha_i u_i$. That is,

$$u = [u]^\alpha, \quad \forall\, u \in \mathbf{F}^n. \qquad (*)$$

Similarly,

$$v = [v]^\beta, \quad \forall\, v \in \mathbf{F}^m. \qquad (**)$$

From Remark 2.5.6,

$$[L_A(u)]^\beta = [L_A]^\beta_\alpha [u]^\alpha.$$

Applying $(*)$ and $(**)$, we get

$$L_A(u) = [L_A]^\beta_\alpha u.$$

Since $L_A(u) = Au$, we get

$$Au = [L_A]^\beta_\alpha u. \qquad \forall\, u \in \mathbf{F}^n$$

Using $u = e_i$ for any $i \in \{1, \ldots, n\}$ shows that the i^{th} column of A is equal to the i^{th} column of $[L_A]^\beta_\alpha$. So, $[L_A]^\beta_\alpha = A$, as desired.

Now, let $T\colon \mathbf{F}^n \to \mathbf{F}^m$ be a linear transformation. From Remark 2.5.6, for any $u \in \mathbf{F}^n$,

$$[T(u)]^\beta = [T]^\beta_\alpha [u]^\alpha.$$

Applying $(*)$ and $(**)$,

$$T(u) = [T]^\beta_\alpha u = L_{[T]^\beta_\alpha}(u). \qquad \forall\, u \in \mathbf{F}^n.$$

Therefore, $T = L_{[T]^\beta_\alpha}$, as desired. □

Lemma 2.5.18. *Let U, V, W, X be vector spaces over a field $\mathbf{F}$. Let $T\colon U \to V$, $S\colon V \to W$, $R\colon W \to X$ be three linear transformations. Then $R(ST) = (RS)T$.*

Proof. We are required to show that, for all $u \in U$, $R(ST)(u) = (RS)T(u)$. We repeatedly apply Definition 2.5.10 as follows.

$$R(ST)(u) = R(ST(u)) = R(S(T(u))) = RS(T(u)) = (RS)T(u).$$

□

Note that Lemma 2.5.18 was proven in a coordinate-free manner. We now combine Lemmas 2.5.17 and 2.5.18 to prove associativity of matrix multiplication, a statement that uses coordinates.

Corollary 2.5.19. *Let A be an $m \times \ell$ matrix, let B be an $n \times m$ matrix, and let C be a $k \times n$ matrix. Then $C(BA) = (CB)A$.*

Proof. From Lemma 2.5.18,

$$L_C(L_B L_A) = (L_C L_B) L_A. \tag{2.5.10}$$

Let $\alpha, \beta, \gamma, \delta$ be the standard bases for $\mathbf{F}^\ell, \mathbf{F}^m, \mathbf{F}^n$ and $\mathbf{F}^k$, respectively. Applying Theorem 2.5.14 twice to the left side of (2.5.10),

$$\begin{aligned} [L_C(L_B L_A)]^\delta_\alpha &= [L_C]^\delta_\gamma [L_B L_A]^\gamma_\alpha = [L_C]^\delta_\gamma ([L_B]^\gamma_\beta [L_A]^\beta_\alpha) \\ &= C(BA) \qquad \text{by Lemma 2.5.17.} \end{aligned} \tag{2.5.11}$$

Applying Theorem 2.5.14 twice to the right side of (2.5.10),

$$\begin{aligned} [(L_C L_B) L_A]^\delta_\alpha &= [L_C L_B]^\delta_\beta [L_A]^\beta_\alpha = ([L_C]^\delta_\gamma [L_B]^\gamma_\beta)[L_A]^\beta_\alpha \\ &= (CB)A \qquad \text{by Lemma 2.5.17.} \end{aligned} \tag{2.5.12}$$

Combining (2.5.10), (2.5.11) and (2.5.12) completes the proof. □

The following facts are proven in a similar manner.

Remark 2.5.20. Let A be an $m \times \ell$ matrix, let B be an $n \times m$ matrix. Then $L_B L_A = L_{BA}$.

Proof. Let α, β, γ be the standard bases for $\mathbf{F}^\ell, \mathbf{F}^m$ and $\mathbf{F}^n$ respectively. Applying Theorem 2.5.14 then Lemma 2.5.17,

$$[L_B L_A]_\alpha^\gamma = [L_B]_\beta^\gamma [L_A]_\alpha^\beta = BA.$$

Taking L of both sides and applying Lemma 2.5.17 to the left side shows that $L_B L_A = L_{BA}$. □

Remark 2.5.21. Let A be an $n \times m$ matrix, let B be an $n \times m$ matrix. Then $L_{A+B} = L_A + L_B$.

Proof. Let α, β be the standard bases for $\mathbf{F}^m$ and $\mathbf{F}^n$, respectively. Applying Lemma 2.5.17,

$$[L_A + L_B]_\alpha^\beta = [L_A]_\alpha^\beta + [L_B]_\alpha^\beta = A + B.$$

Taking L of both sides and applying Lemma 2.5.17 to the left side shows that $L_A + L_B = L_{A+B}$. □

2.6. **Invertibility, Isomorphism.** We now introduce the concept of invertibility. As will become clear, the invertibility of a linear transformation is closely related to our ability to find a "nice" matrix representation of the linear transformation.

Definition 2.6.1 (**Inverse**)**.** Let V, W be vector spaces over a field $\mathbf{F}$. Let $T\colon V \to W$ be a linear transformation. We say that a linear transformation $S\colon W \to V$ is the **inverse** of T if $TS = I_W$ and $ST = I_V$. We say that T is **invertible** if T has an inverse, and we denote the inverse by T^{-1}, so that $TT^{-1} = I_W$ and $T^{-1}T = I_V$.

Remark 2.6.2. If T is the inverse of S, then S is the inverse of T.

If an inverse of T exists, then it is unique, as we now show.

Lemma 2.6.3. *Let V, W be vector spaces over a field $\mathbf{F}$. Let $T\colon V \to W$ be a linear transformation. Let $S\colon W \to V$ be an inverse of T, and let $S'\colon W \to V$ be an inverse of T. Then $S = S'$.*

Proof. Using the definition of inverse,

$$S = SI_W = S(TS') = (ST)S' = I_V S' = S'.$$

□

Lemma 2.6.4. *Let V, W be vector spaces over a field $\mathbf{F}$. Let $T\colon V \to W$ be a linear transformation. If T has an inverse $S\colon W \to V$, then T must be one-to-one and onto.*

Proof. We first show that T is one-to-one. Suppose $v, v' \in V$ satisfy $T(v) = T(v')$. Applying S to both sides, $ST(v) = ST(v')$. That is, $v = v'$, so T is one-to-one, as desired.

We now show that T is onto. Let $w \in W$. We need to find $v \in V$ such that $T(v) = w$. Define $v := Sw$. Then $T(v) = TS(w) = w$, as desired. □

Example 2.6.5. The zero transformation $T\colon \mathbf{R}^2 \to \mathbf{R}^2$ defined by $T = 0$ is not onto, so T is not invertible.

We now prove the converse of Lemma 2.6.4

Lemma 2.6.6. *Let V, W be vector spaces over a field $\mathbf{F}$. Let $T\colon V \to W$ be a linear transformation. Suppose T is one-to-one and onto. Then there exists a linear transformation $S\colon W \to V$ that is the inverse of T.*

Proof. We first have to somehow define a linear transformation $S\colon W \to V$ that inverts T. Given any $w \in W$, since T is bijective, there exists a unique $v \in V$ such that $w = T(v)$. So, define

$$S(w) := v. \qquad (*)$$

Since v uniquely depends on w, the map $S\colon W \to V$ defined in this way is well-defined. We now show that S is linear. Let $w, w' \in W$. Since T is bijective, there exist unique $v, v' \in V$ such that $T(v) = w$ and $T(v') = w'$. In particular, by the definition $(*)$, $S(w) = v$ and $S(w') = v'$. Since T is linear, $T(v + v') = w + w'$. So, by the definition $(*)$, we have $S(w + w') = v + v' = S(w) + S(w')$. Now, let $\alpha \in \mathbf{F}$. Since $T(v) = w$ and T is linear, $T(\alpha v) = \alpha T(v) = \alpha w$. By the definition $(*)$, $S(\alpha w) = \alpha v$. Since $v = S(w)$, we therefore have $S(\alpha w) = \alpha S(w)$, as desired. So, S is linear.

It remains to show that S inverts T. Applying T to both sides of $(*)$, note that $TS(w) = T(v) = w$, so $TS = I_W$. Also, substituting $w = T(v)$ into $(*)$, we get $S(T(v)) = v$, so that $ST = I_V$, as desired. □

Combining Lemmas 2.6.4 and 2.6.6, we see that a linear transformation $T\colon V \to W$ is invertible if and only if T is one-to-one and onto. Invertible linear transformations are also known as **isomorphisms**.

Definition 2.6.7 (**Isomorphism**)**.** Two vector spaces V, W over a field $\mathbf{F}$ are said to be **isomorphic** if there exists an invertible linear transformation $T\colon V \to W$ from one space to the other.

The notion of isomorphism allows us to reason about two vector spaces being the same (if they are isomorphic) or not the same (if they are not isomorphic). Many parts of mathematics, or science more generally, are concerned with classifying things according to whether they are they same or not the same. Within the context of vector spaces, this notion of isomorphism is most appropriate, since it asks for the linear structure of the vector space to be preserved. Within other mathematical contexts, different notions of isomorphism appear, though they all generally ask for the structures at hand to be preserved by a certain map.

Lemma 2.6.8. *Two finite-dimensional vectors spaces V, W over a field $\mathbf{F}$ are isomorphic if and only if* $\dim(V) = \dim(W)$.

Proof. Suppose V, W are isomorphic. Then there exists an invertible linear transformation $T\colon V \to W$. By Lemma 2.6.4, T is one-to-one and onto. In particular, $\text{nullity}(T) = 0$. By the Dimension Theorem (Theorem 2.3.9), $\text{rank}(T) = \dim(V)$. Since T is onto, $\text{rank}(T) = \dim(W)$. Therefore, $\dim(V) = \dim(W)$, as desired.

We now prove the reverse implication. Assume that $\dim(V) = \dim(W) = n$ for some $n \in \mathbf{N}$. Let $\{v_1, \ldots, v_n\}$ be a basis for V, and let $\{w_1, \ldots, w_n\}$ be a basis for W. By Theorem 2.4.4, there exists a linear transformation $T\colon V \to W$ such that $T(v_i) = w_i$ for all $i \in \{1, \ldots, n\}$. By Theorem 2.4.1, $\{w_1, \ldots, w_n\}$ spans $R(T)$. Since $\{w_1, \ldots, w_n\}$ also spans W, we have $R(T) = W$, so that T is onto. By Lemma 2.3.10 (using $\dim(V) = \dim(W)$), T is also one-to-one. So, T is an isomorphism, and V, W are isomorphic, as desired. □

Remark 2.6.9. If V has an ordered basis $\beta = (v_1, \ldots, v_n)$, then the coordinate map $\phi_\beta\colon V \to \mathbf{F}^n$ defined by

$$\phi_\beta(v) := [v]^\beta$$

is a linear transformation. It is also an isomorphism. Note that ϕ_β is one-to-one by Theorem 1.6.11, and ϕ_β is onto since, if we are given the coordinate vector $[v]^\beta = (\alpha_1, \ldots, \alpha_n)$, then $\phi_\beta(\sum_{i=1}^n \alpha_i v_i) = [v]^\beta$. So, ϕ_β is an isomorphism by Lemma 2.6.6. The book calls ϕ_β the **standard representation** of V with respect to β.

If we only care about linear transformations for finite-dimensional vector spaces over $\mathbf{R}$, then Lemma 2.6.8 and Theorem 2.5.14 show that it suffices to discuss real matrices and the vector spaces $\mathbf{R}^n$, $n \in \mathbf{N}$. However, our effort in developing the theory of linear transformations was not a waste of time. For example, the notion of isomorphism from Definition 2.6.7 is not very meaningful for infinite-dimensional vector spaces. For another example, when we introduce norms and inner products, the notion of isomorphism from Definition 2.6.7 becomes less meaningful, and finer properties of linear transformations become more relevant. Nevertheless, we will mostly discuss real matrices and $\mathbf{R}^n$ for the rest of the course.

2.6.1. *Invertibility and Matrices.*

Definition 2.6.10 (**Inverse Matrix**)**.** Let A be an $m \times n$ matrix. We say that A has an **inverse** B if B is an $n \times m$ matrix such that $AB = I_m$ and such that $BA = I_n$. If A has an inverse, we say that A is an **invertible matrix**, and we write $B = A^{-1}$.

We now continue to emphasize the relation between linear transformations and matrices, as in Theorem 2.5.14 and Remark 2.5.6.

Theorem 2.6.11. *Let V, W be vector spaces over a field $\mathbf{F}$. Assume that α is an ordered basis for V with n elements, and assume that β is an ordered basis for W with m elements. Then a linear transformation $T\colon V \to W$ is invertible if and only if $[T]_\alpha^\beta$ is invertible. Also, $[T^{-1}]_\beta^\alpha = ([T]_\alpha^\beta)^{-1}$*

Proof. Suppose $T\colon V \to W$ has an inverse $T^{-1}\colon W \to V$. Then $TT^{-1} = I_W$ and $T^{-1}T = I_V$. So, applying Theorem 2.5.14,

$$[T]_\alpha^\beta [T^{-1}]_\beta^\alpha = [TT^{-1}]_\beta^\beta = [I_W]_\beta^\beta = I_m.$$

$$[T^{-1}]_\beta^\alpha [T]_\alpha^\beta = [T^{-1}T]_\alpha^\alpha = [I_V]_\alpha^\alpha = I_n.$$

So, $[T^{-1}]_\beta^\alpha$ is the inverse of $[T]_\alpha^\beta$, so that $[T]_\alpha^\beta$ is an invertible matrix.

We now prove the reverse implication. Suppose $[T]_\alpha^\beta$ is invertible. Then there exists an $n \times m$ matrix B such that $B[T]_\alpha^\beta = I_n$ and $[T]_\alpha^\beta B = I_m$. Write $\alpha = (v_1, \ldots, v_n)$, $\beta = (w_1, \ldots, w_m)$. We would like to have a linear transformation $S\colon W \to V$ such that $S(w_i) = \sum_{k=1}^n B_{ki} v_k$ for all $i \in \{1, \ldots, m\}$. If such an S exists, then $[S]_\beta^\alpha = B$. Such a linear transformation exists by Theorem 2.4.4. Therefore,

$$[I_V]_\alpha^\alpha = I_n = B[T]_\alpha^\beta = [S]_\beta^\alpha [T]_\alpha^\beta = [ST]_\alpha^\alpha.$$

$$[I_W]_\beta^\beta = I_m = [T]_\alpha^\beta B = [T]_\alpha^\beta [S]_\beta^\alpha = [TS]_\beta^\beta.$$

So, T is invertible, as desired. □

Corollary 2.6.12. *An $m \times n$ matrix A is invertible if and only if the linear transformation $L_A\colon \mathbf{F}^n \to \mathbf{F}^m$ is invertible. Also, $(L_A)^{-1} = L_{A^{-1}}$.*

Proof. Let α be the standard basis for $\mathbf{F}^n$ and let β be the standard basis for $\mathbf{F}^m$. Then

$$[L_A]_\alpha^\beta = A. \qquad (*)$$

So, by Theorem 2.6.11, L_A is invertible if and only if A is invertible. Also, from Theorem 2.6.11,

$$[L_A^{-1}]_\beta^\alpha = ([L_A]_\alpha^\beta)^{-1} = A^{-1} = [L_{A^{-1}}]_\beta^\alpha \qquad \text{, by } (*).$$

Therefore, $L_A^{-1} = L_{A^{-1}}$. □

Corollary 2.6.13. *Let A be an $m \times n$ matrix. If A is invertible, then $m = n$.*

Proof. Apply Corollary 2.6.12 and Lemma 2.6.8. □

Unfortunately, not all matrices are invertible. For example, the zero matrix is not invertible.

2.7. **Change of Coordinates.** Suppose we have two finite ordered bases β, β' for the same vector space V. Let $v \in V$. We would like a way to relate $[v]^\beta$ to $[v]^{\beta'}$. Using Remark 2.5.6 and that $I_V v = v$ for all $v \in V$, we have

$$[I_V]_\beta^{\beta'}[v]^\beta = [v]^{\beta'}.$$

That is, to relate $[v]^\beta$ to $[v]^{\beta'}$, it suffices to compute $[I_V]_\beta^{\beta'}$

Example 2.7.1. Let $\beta = ((2,0),(1,-1))$, and let $\beta' = ((0,1),(2,1))$ be two ordered bases of $\mathbf{R}^2$. Then

$$I_V(2,0) = (2,0) = -1(0,1) + 1(2,1).$$
$$I_V(1,-1) = (1,-1) = -(3/2)(0,1) + (1/2)(2,1).$$

So

$$[I_V]_\beta^{\beta'} = \begin{pmatrix} -1 & -3/2 \\ 1 & 1/2 \end{pmatrix}.$$

So, we can verify that $[I_V]_\beta^{\beta'}[v]^\beta = [v]^{\beta'}$. For example, choosing $v = (3,2)$, note that

$$[v]^\beta = \begin{pmatrix} 5/2 \\ -2 \end{pmatrix}, \qquad [v]^{\beta'} = \begin{pmatrix} 1/2 \\ 3/2 \end{pmatrix}, \qquad \begin{pmatrix} -1 & -3/2 \\ 1 & 1/2 \end{pmatrix}\begin{pmatrix} 5/2 \\ -2 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 3/2 \end{pmatrix}.$$

Similarly, note that $[I_V]_{\beta'}^\beta$ is the inverse of $[I_V]_\beta^{\beta'}$, so

$$[I_V]_{\beta'}^\beta = \begin{pmatrix} 1/2 & 3/2 \\ -1 & -1 \end{pmatrix}.$$

Exercise 2.7.2. Show that $[I_V]_\beta^{\beta'}$ is invertible, with inverse $[I_V]_{\beta'}^\beta$.

Lemma 2.7.3. *Let V be a finite-dimensional vector space over a field $\mathbf{F}$. Let β, β' be two bases for V. Let $T\colon V \to V$ be a linear transformation. Define $Q := [I_V]_\beta^{\beta'}$. (From Theorem 2.6.11, Q is invertible.) Then $[T]_\beta^\beta$ and $[T]_{\beta'}^{\beta'}$ satisfy the following relation*

$$[T]_{\beta'}^{\beta'} = Q[T]_\beta^\beta Q^{-1}.$$

Proof. We first write $T = I_V T I_V$. Taking the matrix representation of both sides and then applying Theorem 2.5.14,

$$[T]_{\beta'}^{\beta'} = [I_V T I_V]_{\beta'}^{\beta'} = [I_V]_{\beta}^{\beta'}[TI_V]_{\beta'}^{\beta} = [I_V]_{\beta}^{\beta'}[T]_{\beta}^{\beta}[I_V]_{\beta'}^{\beta}.$$

From Theorem 2.6.11, $[I_V]_{\beta'}^{\beta} = ([I_V]_{\beta}^{\beta'})^{-1}$, completing the proof. □

Definition 2.7.4 (**Similarity**)**.** Two $n \times n$ matrices A, B are said to be **similar** if there exists an invertible $n \times n$ matrix Q such that $A = QBQ^{-1}$.

Remark 2.7.5. In the context of Lemma 2.7.3, $[T]_{\beta'}^{\beta'}$ is similar to $[T]_{\beta}^{\beta}$.

3. Row Operations, The Determinant

3.2. **Row Operations.** We begin our discussion of row operations on matrices with some examples.

Example 3.2.1 (**Type 1: Interchange two Rows**)**.** For example, we can swap the first and third rows of the matrix

$$\begin{pmatrix} 1 & 2 \\ 3 & 5 \\ 0 & 8 \end{pmatrix}$$

to get

$$\begin{pmatrix} 0 & 8 \\ 3 & 5 \\ 1 & 2 \end{pmatrix}.$$

Define

$$E := \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Note that

$$E \begin{pmatrix} 1 & 2 \\ 3 & 5 \\ 0 & 8 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 5 \\ 0 & 8 \end{pmatrix} = \begin{pmatrix} 0 & 8 \\ 3 & 5 \\ 1 & 2 \end{pmatrix}.$$

Remark 3.2.2. E as defined above is invertible. In fact, $E = E^{-1}$. In general, if E is the $n \times n$ matrix that swaps two rows of an $n \times n$ matrix A, then EA is A with those two rows swapped. So $EEA = A$ for all $n \times n$ matrices A, so $EE = I_n$, i.e. E is invertible.

Example 3.2.3 (**Type 2: Multiply a row by a nonzero scalar**)**.** For example, let's multiply the second row of the following matrix by 2.

$$\begin{pmatrix} 1 & 2 \\ 3 & 5 \\ 0 & 8 \end{pmatrix}.$$

We then get

$$\begin{pmatrix} 1 & 2 \\ 6 & 10 \\ 0 & 8 \end{pmatrix}.$$

Define

$$E := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Note that

$$E \begin{pmatrix} 1 & 2 \\ 3 & 5 \\ 0 & 8 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 5 \\ 0 & 8 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 6 & 10 \\ 0 & 8 \end{pmatrix}$$

Remark 3.2.4. E as defined above has inverse

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

In general, suppose E corresponds to multiplying the i^{th} row of a given matrix by $\alpha \in \mathbf{F}$, $\alpha \neq 0$. Then E is a matrix with ones on the diagonal, except for the i^{th} entry on the diagonal, which is α. And all other entries of E are zero. Then, we see that E^{-1} exists and is a matrix with ones on the diagonal, except for the i^{th} entry on the diagonal, which is α^{-1}. And all other entries of E^{-1} are zero. In particular, E is invertible.

Example 3.2.5 (**Adding one row to another**)**.** Let's add two copies of the first row of the following matrix to the third row.

$$\begin{pmatrix} 1 & 2 \\ 3 & 5 \\ 0 & 8 \end{pmatrix}.$$

We then get

$$\begin{pmatrix} 1 & 2 \\ 3 & 5 \\ 2 & 12 \end{pmatrix}.$$

Define

$$E := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix}.$$

Note that

$$E \begin{pmatrix} 1 & 2 \\ 3 & 5 \\ 0 & 8 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 5 \\ 0 & 8 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 5 \\ 2 & 12 \end{pmatrix}.$$

Remark 3.2.6. E as defined above has inverse

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}.$$

That is, adding 2 copies of row one to row three is inverted by adding -2 copies of row one to row three. In a similar way, a general row addition operator is seen to be invertible.

Remark 3.2.7 (**Summary of Row Operations**)**.** The three row operations (Type 1, Type 2, and Type 3) are all invertible.

Remark 3.2.8 (**Solving Systems of Linear Equations**). Let A be an $m \times n$ matrix, let $x \in \mathbf{R}^n$ be a variable vector, and let $b \in \mathbf{R}^m$ be a known vector. Consider the system of linear equations

$$Ax = b.$$

Let E be any elementary row operation. Since E is invertible, finding a solution x to the system $Ax = b$ is equivalent to finding the solution x to the system $EAx = Eb$. By applying many elementary row operations, you have seen in a previous course how to solve the system $Ax = b$. That is, you continue to apply elementary row operations $E_1, \ldots, E_k$ such that $E_1 \cdots E_k A$ in in **row-echelon** form, and you then solve $E_1 \cdots E_k Ax = E_1 \cdots E_k b$. A matrix B is in row-echelon form if each row is either zero, or its left-most nonzero entry is 1, with zeros below the 1.

Remark 3.2.9 (**Inverting a Matrix**). Let A be an invertible $n \times n$ matrix. You learned in a previous course an algorithm for inverting A using elementary row operations. Below, we will prove that this algorithm works.

Remark 3.2.10 (**Column Operations**). In the above discussion, we could have also used column operations instead of row operations. Column operations would then correspond to multiplying the matrices E on the right side, rather than the left side. The invertibility of column operations would therefore still hold.

3.3. **Rank of a Matrix.** Let $T\colon V \to W$ be a linear transformation between two vector spaces. Recall that the rank of T, denoted by $\operatorname{rank}(T)$, is the dimension of $R(T)$, the range of T.

Lemma 3.3.1. *Let V, W be finite-dimensional vector spaces over a field $\mathbf{F}$. Assume that $\dim(V) = \dim(W) = n$. Let $T\colon V \to W$ be a linear transformation. Then T is invertible if and only if T has rank n.*

Proof. Suppose T is invertible. Then T is one-to-one. By the Dimension Theorem (Theorem 2.3.9), T has rank n.

Now, suppose T has rank n. Then, by the Dimension Theorem, $N(T) = \{0\}$, so T is one-to-one. Also, $R(T)$ is again a subspace of W of the same dimension as W, so we must have $R(T) = W$, so T is onto. Since T is both one-to-one and onto, T is invertible. □

Lemma 3.3.2. *Let V, W be finite-dimensional vector spaces over a field $\mathbf{F}$. Let $T\colon V \to W$ be an invertible linear transformation. Let $U \subseteq V$ be a subspace. Then $\dim(U) = \dim(T(U))$.*

Proof. Since $U \subseteq V$ is a subspace, U is a vector space. So, for each $u \in U$, define the map $T_U\colon U \to W$ by

$$T_U(u) := T(u).$$

Since T is linear, T_U is linear. Since T is one-to-one, T_U is one-to-one, so $N(T_U) = \{0\}$. So, the Dimension Theorem (Theorem 2.3.9) implies that $\dim R(T_U) = \dim(U)$. Since $R(T_U) = T_U(U) = T(U)$, we are done. □

Lemma 3.3.3 (**Isomorphisms Preserve Rank**). *Let U, V, W, X be vector spaces over a field $\mathbf{F}$. Let $T\colon V \to W$ be a linear transformation. Let $S\colon U \to V$ be an invertible linear transformation, and let $P\colon W \to X$ be an invertible linear transformation. Then*

$$\operatorname{rank}(T) = \operatorname{rank}(PT) = \operatorname{rank}(TS) = \operatorname{rank}(PTS).$$

Proof. We begin with the first equality. By the definition of range, $R(T) = T(V)$, and $R(PT) = PT(V)$. So,

$$R(PT) = PT(V) = P(T(V)) = P(R(T)).$$

So, $\text{rank}(PT) = \dim(P(R(T)))$. Since P is invertible, $\dim(P(R(T))) = \dim(R(T))$ by Lemma 3.3.2. So, $\text{rank}(PT) = \text{rank(T)}$.

We now prove that $\text{rank}(T) = \text{rank}(TS)$. Since $S\colon U \to V$ is invertible, S is onto. So, $S(U) = V$. By the definition of range,

$$R(TS) = TS(U) = T(S(U) = T(V).$$

So, $R(TS) = T(V) = R(T)$, so $\text{rank}(T) = \text{rank}(TS)$.

Finally, the equality $\text{rank}(PTS) = \text{rank}(TS)$ follows by applying the first equality to $T' := TS$. □

Definition 3.3.4 (**Rank of a Matrix**)**.** Let A be a matrix. Then the **rank** of A is defined as $\text{rank}(L_A)$.

Lemma 3.3.5. *The rank of a matrix A is equal to the dimension of the space spanned by the columns of A.*

Proof. Suppose A is an $m \times n$ matrix. Let $(e_1, \ldots, e_n)$ be the standard basis of $\mathbf{F}^n$. Since this basis spans $\mathbf{F}^n$, the vectors $\{L_A(e_1), \ldots, L_A(e_n)\}$ span $R(L_A)$ by Theorem 2.4.1. But for each $i \in \{1, \ldots, n\}$, $L_A(e_i)$ is the i^{th} column of A. □

Remark 3.3.6. Suppose V and W are finite dimensional vector spaces. Let α, γ be ordered bases for V and let β, δ be ordered bases for W. Let $T\colon V \to W$ be a linear transformation. Recall that any two matrix representations $[T]^\beta_\alpha$ and $[T]^\delta_\gamma$ are related by the identity $[T]^\delta_\gamma = [I_W]^\delta_\beta [T]^\beta_\alpha [I_V]^\alpha_\gamma$. Also, two vector spaces of the same dimension are isomorphic. So, to compute the rank of T, it suffices to find any matrix representation A of T, and then to compute the rank of A. By Lemma 2.5.18, isomorphisms preserve rank, so any matrix representation A suffices. And we can compute the rank of A by row-reducing it into row-echelon form, and then applying the following lemma.

Lemma 3.3.7. *Let A be a matrix in row-echelon form. Then the rank of A is equal to the number of nonzero rows of A.*

Proof. Since A is in row-echelon form, after permuting the rows of A, there exists a positive integer k such that, each of the first k rows of A has one nonzero entry, while all subsequent rows of A are zero. So, the span of the columns of A are contained in the k-dimensional subspace

$$\left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{pmatrix} : x_1, \ldots, x_k \in \mathbf{F} \right\}. \qquad (*)$$

So, by Lemma 3.3.5, $\text{rank}(A) \leq k$. We now show that in fact $\text{rank}(A) = k$, as desired.

By Lemma 3.3.5, it suffices to show that the span of the columns of A contains the subspace $(*)$. To show this, let v be in the subspace $(*)$. Then, as a column vector, we

have $v = (v_1, \ldots, v_k, 0, \ldots, 0)$, $v_i \in \mathbf{F}$ for all $i \in \{1, \ldots, k\}$. Consider the i^{th} row of A where $1 \leq i \leq k$. Since A is in row-echelon form, the i^{th} row first has several zeros, then a 1, then other entries afterwards. So, for each $i \in \{1, \ldots, k\}$, there exists $j(i)$ such that the $j(i)^{th}$ column of A has a 1 in the i^{th} row, and then zeros below that. So, beginning with $i = k$, we can subtract v_k copies of the $j(k)^{th}$ column of A from v, giving a vector with only $(k-1)$ nonzero entries. Then, setting $i = k-1$, we can subtract copies of the $j(k-1)^{st}$ column of A to get a vector with only $(k-2)$ nonzero entries. We continue in this way, and eventually we have eliminated all nonzero entries of v. That is, we have found an expression for v in terms of the columns $j(1), \ldots, j(k)$ of A. So, $\text{rank}(A) = k$, as desired.

□

Theorem 3.3.8. *Let A be an $m \times n$ matrix of rank r. Then, there exist a finite number of elementary row and column operations which, when applied to A, produce the matrix*

$$\begin{pmatrix} I_{r\times r} & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{pmatrix}.$$

Proof. We first use row reduction to put A into row-echelon form. So, after this row reduction, the first r rows of A have some zeros, and then a 1 with zeros below this 1. And the remaining $m - r$ rows are all zero. (In case $r = 0$, then we are done, so we may assume that $r > 0$.) Now, the first row of A has some zeros, then a 1 with zeros below this 1. So, by adding copies of the column that contains the entry 1 to each column to the right, the remaining entries of the first row can be made to be zero. And we still keep our matrix in row-echelon form. Now, the second row of A has some zeros, then a 1 with zeros above and below this 1. So, by adding copies of the column that contains this entry 1 to each column to the right, the remaining entries of the second row can be made to be zero. And once again, our matrix is still in row-echelon form. We then continue this procedure. The first r rows then each have exactly one entry of 1, and all remaining entries in the matrix are zero. By swapping columns as needed, A is then put into the required form, as desired. □

Corollary 3.3.9 (**A Factorization Theorem**). *Let A be an $m \times n$ matrix of rank r. Then, there exists an $m \times m$ matrix B and an $n \times n$ matrix C such that B is the product of a finite number of elementary row operations, C is the product of a finite number of elementary column operations, and such that*

$$A = B \begin{pmatrix} I_{r\times r} & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{pmatrix} C.$$

Proof. Let A be an $m \times n$ matrix of rank r. From Theorem 3.3.8, there exist a finite number of elementary row operations $E_1, \ldots, E_j$ and elementary column operations $F_1, \ldots, F_k$ such that

$$E_1 \cdots E_j A F_1 \cdots F_k = \begin{pmatrix} I_{r\times r} & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{pmatrix}. \qquad (*)$$

From Remarks 3.2.7 and 3.2.10, the matrices $E_1, \ldots, E_j$ and $F_1, \ldots, F_k$ are invertible, with inverses that are also elementary row and column operations, respectively. So, multiplying on the left of each side of $(*)$ by $B := E_j^{-1} \cdots E_1^{-1}$, and then multiplying on the right of each side of $(*)$ by $C := F_k^{-1} \cdots F_1^{-1}$, we deduce the theorem. □

Lemma 3.3.10. *Let A be an $m \times n$ matrix. Let B be an $m \times m$ invertible matrix, and let C be an $n \times n$ invertible matrix. Then*

$$\operatorname{rank}(A) = \operatorname{rank}(BA) = \operatorname{rank}(AC) = \operatorname{rank}(BAC).$$

Proof. Since B is invertible, L_B is invertible with inverse $L_{B^{-1}}$, by Corollary 2.6.12. So, applying Remark 2.5.20 and Lemma 3.3.3,

$$\operatorname{rank}(L_A) = \operatorname{rank}(L_{BA}) = \operatorname{rank}(L_{AC}) = \operatorname{rank}(L_{BAC}).$$

Definition 3.3.4 then completes the proof. □

Definition 3.3.11 (**Transpose**)**.** Let A be an $m \times n$ matrix with entries A_{ij}, $1 \leq i \leq m$, $1 \leq j \leq n$. Then the **transpose** A^t of A is defined to be the $n \times m$ matrix with entries $(A^t)_{ij} := A_{ji}$, $1 \leq i \leq n$, $1 \leq j \leq m$.

Exercise 3.3.12. Let A be an $m \times n$ matrix. Let B be an $\ell \times m$ matrix. Show that $(BA)^t = A^t B^t$.

Remark 3.3.13. If A is an $n \times n$ invertible matrix, then $I_n^t = (AA^{-1})^t = (A^{-1})^t A^t$, so A^t is also invertible.

Lemma 3.3.14. *Let A be an $m \times n$ matrix with rank r. Then A^t also has rank r.*

Proof. From Theorem 3.3.9, there exists an invertible $m \times m$ matrix B and an invertible $n \times n$ matrix C such that

$$A = B \begin{pmatrix} I_{r\times r} & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{pmatrix} C.$$

Taking the transpose of both sides and applying Exercise 3.3.12,

$$A^t = C^t \begin{pmatrix} I_{r\times r} & 0_{r\times(m-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(m-r)} \end{pmatrix} B^t.$$

From Remark 3.3.13, C^t and B^t are invertible. So, Lemma 3.3.10 implies that A^t has rank r. □

Corollary 3.3.15. *The rank of a matrix is equal to the dimension of the span of its rows.*

Proof. Apply Lemma 3.3.5 and Lemma 3.3.14. □

Lemma 3.3.16. *Let V be an n-dimensional vector space, and let W be an m-dimensional vector space. Let $T\colon V \to W$ be a linear transformation. Let α, β be finite bases for V, W respectively. Then* $\operatorname{rank}(T) = \operatorname{rank}([T]_\alpha^\beta)$.

Proof. Let $v \in V, w \in W$. The coordinate maps $\phi_\alpha\colon V \to \mathbf{F}^n$ and $\phi_\beta\colon W \to \mathbf{F}^m$ defined by $\phi_\alpha(v) := [v]^\alpha$, $\phi_\beta(w) := [w]^\beta$ are isomorphisms. Also, the map $L_{[T]_\alpha^\beta}\colon \mathbf{F}^n \to \mathbf{F}^m$ is a linear transformation. Beginning with the identity

$$[T(v)]^\beta = [T]_\alpha^\beta [v]^\alpha,$$

we then rewrite this as

$$\phi_\beta(T(v)) = L_{[T]_\alpha^\beta} \phi_\alpha(v).$$

Since this equality holds for all $v \in V$, we therefore have

$$\phi_\beta T = L_{[T]_\alpha^\beta} \phi_\alpha.$$

Since ϕ_β is invertible, we then get

$$T = \phi_\beta^{-1} L_{[T]_\alpha^\beta} \phi_\alpha.$$

So, applying Lemma 3.3.10 and Definition 3.3.4,

$$\operatorname{rank}(T) = \operatorname{rank}(\phi_\beta^{-1} L_{[T]_\alpha^\beta} \phi_\alpha) = \operatorname{rank}(L_{[T]_\alpha^\beta}) = \operatorname{rank}([T]_\alpha^\beta).$$

□

Exercise 3.3.17. Show that an $m \times n$ matrix has rank at most $\min(m, n)$.

3.3.1. *Inverting a Matrix.*

Lemma 3.3.18. *Let A be an $n \times n$ matrix. Then A is invertible if and only if it is the product of elementary row and column operations.*

Proof. Suppose A is a product of elementary row and column operation matrices. From Remarks 3.2.7 and 3.2.10, A is a product of invertible matrices, so A is invertible.

Now, suppose A is invertible. Then $L_A \colon \mathbf{F}^n \to \mathbf{F}^n$ is invertible (with inverse $L_{A^{-1}}$). In particular, L_A is onto, so $\operatorname{rank}(L_A) = n$. By Definition 3.3.4, $\operatorname{rank}(A) = n$. Applying our Factorization Theorem (Theorem 3.3.9), there exists a finite number of elementary row operations $E_1, \ldots, E_j$ and elementary column operations $F_1, \ldots, F_k$ such that $A = E_1 \cdots E_j F_1 \cdots F_k$, as desired. □

Remark 3.3.19. Suppose A is an invertible matrix, and we have elementary row operations $E_1, \ldots, E_j$ such that

$$E_1 \cdots E_j A = I_n.$$

Multiplying both sides by A^{-1} on the right,

$$E_1 \cdots E_j I_n = A^{-1}.$$

So, to compute A^{-1} from A, it suffices to find row operations that turn A into the identity. And we then apply these operations to I_n to give A^{-1}. This is the algorithm for computing the inverse A^{-1} that you learned in a previous class.

3.4. **The Determinant.** There are a lot of nice things to say about the determinant, but we do not have sufficient time to discuss these things. We will therefore just state some facts about the determinant without proof, and then prove other things as consequences of these preliminary facts.

Let $A \in \mathbf{F}$. Then $\det(A) := A$.

Let A be a 2×2 matrix. That is, there exist $a, b, c, d \in \mathbf{F}$ such that

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

We then define $\det(A)$ so that

$$\det(A) := \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

Let A be a 3×3 matrix. That is, there exist $a, b, c, d, e, f, g, h, i \in \mathbf{F}$ such that

$$A = \det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}.$$

We now define $\det(A)$ inductively so that

$$\det(A) := a \det \begin{pmatrix} e & f \\ h & i \end{pmatrix} - b \det \begin{pmatrix} d & f \\ g & i \end{pmatrix} + c \det \begin{pmatrix} d & e \\ g & h \end{pmatrix}.$$

Definition 3.4.1. More generally, if A is an $n \times n$ matrix, then for each $i, j \in \{1, \ldots, n\}$, let $\overline{A}_{ij}$ denote the $(n-1) \times (n-1)$ matrix formed by removing the i^{th} row and j^{th} column from A. Then, for any $i \in \{1, \ldots, n\}$, define

$$\det(A) := \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \det(\overline{A}_{ij})$$

If A has columns $v_1, \ldots, v_n$, we write $\det(A) = \det(v_1, \ldots, v_n)$ to emphasize that the determinant is a function of the columns of A.

Remark 3.4.2 (**Properties of the Determinant**). Let $v_1, \ldots, v_n \in \mathbf{F}^n$.

(a) For all $\alpha \in \mathbf{F}$, for all $w \in \mathbf{F}^n$, for all $i \in \{1, \ldots, n\}$

$$\begin{aligned} &\det(v_1, \ldots, v_{i-1}, v_i + \alpha w, v_{i+1}, \ldots, v_n) \\ &\quad = \det(v_1, \ldots, v_n) + \alpha \det(v_1, \ldots, v_{i-1}, w, v_{i+1}, \ldots, v_n). \qquad \text{(Multilinear)} \end{aligned}$$

(b) For all $i, j \in \{1, \ldots, n\}$ with $i \neq j$,

$$\det(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) = -\det(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n). \qquad \text{(Alternating)}$$

(c) $\det(I_n) = 1$. (Normalized)

(d) For all $n \times n$ matrices A, B, we have $\det(AB) = \det(A)\det(B)$.

(e) For all $n \times n$ matrices A, we have $\det(A) = \det(A^t)$.

Theorem 3.4.3. *Suppose we have two functions F, G that map $v_1, \ldots, v_n \in \mathbf{F}^n$ to $\mathbf{F}$, both satisfying properties $(a), (b)$ and (c) above. Then $F = G$.*

Proof. Define $D(v_1, \ldots, v_n) := F(v_1, \ldots, v_n) - G(v_1, \ldots, v_n)$. We will show that $D = 0$. Since F, G both satisfy properties $(a), (b)$, D satisfies properties $(a), (b)$. Since F, G both satisfy property (c), we have $D(I_n) = 0$. Since D satisfies property (b) and $D(e_1, \ldots, e_n) = 0$, if $(e_{j(1)}, \ldots, e_{j(n)})$ is any permutation of the standard basis $(e_1, \ldots, e_n)$, we have

$$D(e_{j(1)}, \ldots, e_{j(n)}) = 0. \qquad (*)$$

Let $v_i \in \mathbf{F}^n$, and write $v_i = \sum_{j=1}^{n} \alpha_{ij} e_j$, $\alpha_{ik} \in \mathbf{F}$ for all $i, j \in \{1, \ldots, n\}$. Repeatedly applying property (a),

$$\begin{aligned} D(v_1, \ldots, v_n) &= D(\sum_{j=1}^{n} \alpha_{1j} e_j, v_2, \ldots, v_n) = \sum_{j=1}^{n} \alpha_{1j} D(e_j, v_2, \ldots, v_n) \\ &= \sum_{j_1=1}^{n} \alpha_{1j_1} D(e_{j_1}, \sum_{j=1}^{n} \alpha_{2j} e_j, \ldots, v_n) = \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \alpha_{1j_1} \alpha_{2j_2} D(e_{j_1}, e_{j_2}, \ldots, v_n) \\ &= \cdots = \sum_{j_1=1}^{n} \cdots \sum_{j_n=1}^{n} \alpha_{1j_1} \cdots \alpha_{nj_n} D(e_{j_1}, \ldots, e_{j_n}). \end{aligned}$$

And the final quantity is zero, by $(*)$, as desired. $\square$

Theorem 3.4.3 can be used to show that various different definitions of the determinant all agree. Given some formula that should be equal to the determinant, it suffices to prove that this formula satisfies properties $(a), (b)$ and (c). For example, consider the following determinant formula you learned in Calc 3, for vectors $v_1, v_2, v_3 \in \mathbf{R}^3$:

$$\det(v_1, v_2, v_3) = v_1 \cdot (v_2 \times v_3).$$

Here $\cdot$ denotes the dot product, and $\times$ denotes the cross product. You could write the right side in coordinates to verify that it agrees with the left side. Or, you could verify that the right side satisfies properties $(a), (b)$ and (c), and then apply Theorem 3.4.3, instead giving a coordinate-free proof of the desired identity.

As another application of Theorem 3.4.3, we can show that property (d) of Remark 3.4.2 holds.

Theorem 3.4.4. *Assume that the determinant function satisfies properties $(a), (b)$ and (c) from Remark 3.4.2. Then the determinant function satisfies property (d). For all $n \times n$ matrices A, B, we have* $\det(AB) = \det(A)\det(B)$.

Proof. Suppose $\det(A) \neq 0$. For $v_1, \ldots, v_n \in \mathbf{F}^n$, define

$$F(v_1, \ldots, v_n) := \det(Av_1, \ldots, Av_n)/\det(A).$$

Note that F then satisfies properties $(a), (b)$ and (c). So, we have by Theorem 3.4.3 that $F(B) = \det(AB)/\det(A) = \det(B)$. So, $\det(AB) = \det(A)\det(B)$, as desired.

In the case $\det(A) = 0$, define

$$F(v_1, \ldots, v_n) := \det(v_1, \ldots, v_n) + \det(Av_1, \ldots, Av_n).$$

Once again, F satisfies properties $(a), (b)$ and (c), so $F(B) = \det(B) = \det(B) + \det(AB)$. So, $\det(AB) = 0 = \det(A)\det(B)$. In any case, $\det(AB) = \det(A)\det(B)$, as desired. □

Theorem 3.4.5. *Let A be an $n \times n$ matrix. Then A is invertible if and only if* $\det(A) \neq 0$. *If A is invertible, then* $\det(A^{-1}) = (\det(A))^{-1}$.

Proof. Suppose A has rank r. From the Factorization Theorem (Theorem 3.3.9), A is the product of elementary row and column operations, and also a diagonal matrix D with r ones on the diagonal. If $r < n$, then $\det(D) = 0$, so $\det(A) = 0$ as well from property (d) of Remark 3.4.2. We have shown that, if A has rank less than n, then $\det(A) = 0$. Taking the contrapositive, if $\det(A) \neq 0$, then A has rank n. From Lemma 3.3.1, A is invertible if and only if A has rank n. So, if $\det(A) \neq 0$, then A is invertible.

We now prove the converse. Suppose A is invertible. From property (d) of Remark 3.4.2, $1 = \det(I_n) = \det(AA^{-1}) = \det(A)\det(A^{-1})$. So, $\det(A)$ must be nonzero. □

Corollary 3.4.6. *For any $n \times n$ matrix A,* $\det(A) = \det(A^t)$.

Proof. Suppose A has rank r. From the Factorization Theorem (Theorem 3.3.9), there exist elementary row operations $E_1, \ldots, E_j$ and elementary column operations $F_1, \ldots, F_k$, and there exists a diagonal matrix D with r ones on the diagonal such that

$$A = E_1 \cdots E_j D F_1 \cdots F_k \qquad (*).$$

Taking the transpose of $(*)$,

$$A^t = F_k^t \cdots F_1^t D E_j^t \cdots E_1^t. \qquad (**)$$

From Theorem 3.4.4 applied to $(*)$

$$\det(A) = \det(E_1)\cdots\det(E_j)\det(D)\det(F_1)\cdots\det(F_k).$$

By checking Type 1, 2 and 3 matrices from Examples 3.2.1, 3.2.3 and 3.2.5 directly, we see that $\det(E) = \det(E^t)$ for any elementary row or column operation E. So, applying Theorem 3.4.4 to $(**)$,

$$\begin{aligned}\det(A^t) &= \det(F_k^t)\cdots\det(F_1^t)\det(D)\det(E_j^t)\cdots\det(E_1^t)\\ &= \det(E_1)\cdots\det(E_j)\det(D)\det(F_1)\cdots\det(F_k) = \det(A).\end{aligned}$$

□

4. Eigenvalues, Eigenvectors, Diagonalization

Exercise 4.0.1. Let A be an $n \times n$ matrix with entries A_{ij}, $i, j \in \{1, \ldots, n\}$, and let S_n denote the set of all permutations on n elements. For $\sigma \in S_n$, let $\text{sign}(\sigma) := (-1)^N$, where σ can be written as a composition of N transpositions. Then

$$\det(A) = \sum_{\sigma \in S_n} \text{sign}(\sigma) \prod_{i=1}^{n} A_{i\sigma(i)}.$$

4.2. **Diagonal Matrices.** So far, we should have a reasonably good understanding of linear transformations, matrices, rank and invertibility. However, given a matrix, we don't yet have a good understanding of how to "simplify" this matrix. In mathematics and science, the general goal is to take some complicated and make it simpler. In the context of linear algebra, this paradigm becomes: try to find a particular basis such that a linear transformation has a diagonal matrix representation. (After all, diagonal matrices are among the simplest matrices.) We now attempt to realize this goal within our discussion of eigenvectors and diagonalization.

Definition 4.2.1 (**Diagonal Matrix**)**.** An $n \times n$ matrix A with entries A_{ij}, $i, j \in \{1, \ldots, n\}$ is said to be **diagonal** if $A_{ij} = 0$ whenever $i \neq j$, $i, j \in \{1, \ldots, n\}$. If A is diagonal, we denote the matrix A by $\text{diag}(A_{11}, A_{22}, \ldots, A_{nn})$.

Lemma 4.2.2. *The rank of a diagonal matrix is equal to the number of its nonzero entries.*

4.3. **Eigenvectors and Eigenvalues.**

Definition 4.3.1 (**Eigenvector and Eigenvalue**)**.** Let V be a vector space over a field $\mathbf{F}$. Let $T\colon V \to V$ be a linear transformation. An **eigenvector** of T is a nonzero vector $v \in V$ such that, there exists $\lambda \in \mathbf{F}$ with $T(v) = \lambda v$. The scalar λ is then referred to as the **eigenvalue** of the eigenvector v.

Remark 4.3.2. The word "eigen" is German for "self." The equation $T(v) = \lambda v$ is self-referential in v, which explains the etymology here.

Example 4.3.3. If A is diagonal with $A = \text{diag}(A_{11}, \ldots, A_{nn})$, then L_A has eigenvectors $(e_1, \ldots, e_n)$ with eigenvalues $(A_{11}, \ldots, A_{nn})$.

Example 4.3.4. If T is the identity transformation, then every vector is an eigenvector with eigenvalue 1.

Example 4.3.5. If $T\colon V \to V$ has $v \in N(T)$ with $v \neq 0$, then v is an eigenvector of T with eigenvalue zero.

Example 4.3.6. Define $T\colon C^\infty(\mathbf{R}) \to C^\infty(\mathbf{R})$ by $T(f) := -f''$. For any $y \in \mathbf{R}$, the function $f(x) := e^{ixy}$ satisfies $Tf(x) = f''(x) = y^2 f(x)$. So, for any $y \in \mathbf{R}$, e^{ixy} is an eigenfunction of T with eigenvalue y^2.

Definition 4.3.7 (**Eigenspace**)**.** Let V be a vector space over a field $\mathbf{F}$. Let $T\colon V \to V$ be a linear transformation. Let $\lambda \in \mathbf{F}$. The **eigenspace** of λ is the set of all $v \in V$ (including zero) such that $T(v) = \lambda v$.

Remark 4.3.8. Given $\lambda \in \mathbf{F}$, the set of v such that $T(v) = \lambda v$ is the same as $N(T - \lambda I_V)$. In particular, an eigenspace is a subspace of V. And $N(T - \lambda I_V)$ is nonzero if and only if $T - \lambda I_V$ is not one-to-one.

Lemma 4.3.9 (**An Eigenvector Basis Diagonalizes** T)**.** *Let V be an n-dimensional vector space over a field $\mathbf{F}$, and let $T\colon V \to V$ be a linear transformation. Suppose V has an ordered basis $\beta := (v_1, \ldots, v_n)$. Then v_i is an eigenvector of T with eigenvalue $\lambda_i \in \mathbf{F}$, for all $i \in \{1, \ldots, n\}$, if and only if the matrix $[T]_\beta^\beta$ is diagonal with $[T]_\beta^\beta = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$.*

Proof. We begin with the forward implication. Let $i \in \{1, \ldots, n\}$. Suppose $T(v_i) = \lambda_i v_i$, $[T(v_i)]^\beta$ is a column vector whose i^{th} entry is λ_i, with all other entries zero. Since $[T]_\beta^\beta = ([T(v_1)]^\beta, \ldots, [T(v_n)]^\beta)$, we conclude that $[T]_\beta^\beta = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$.

Conversely, suppose $[T]_\beta^\beta = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Since $[T]_\beta^\beta = ([T(v_1)]^\beta, \ldots, [T(v_n)]^\beta)$, we conclude that $T(v_i) = \lambda_i v_i$ for all $i \in \{1, \ldots, n\}$, so that v_i is an eigenvector of T with eigenvalue λ_i, for all $i \in \{1, \ldots, n\}$. □

Definition 4.3.10 (**Diagonalizable**)**.** A linear transformation $T\colon V \to V$ is said to be **diagonalizable** if there exists an ordered basis β of V such the matrix $[T]_\beta^\beta$ is diagonal.

Remark 4.3.11. From Lemma 4.3.9, T is diagonalizable if and only if it has a basis consisting of eigenvectors of T.

Example 4.3.12. Let $T\colon \mathbf{R}^2 \to \mathbf{R}^2$ denote reflection across the line ℓ which passes through the origin and $(1, 2)$. Then $T(1, 2) = (1, 2)$, and $T(2, -1) = -(2, -1)$, so we have two eigenvectors of T with eigenvalues 1 and -1 respectively. The vectors $((1, 2), (2, -1))$ are independent, so they form a basis of $\mathbf{R}^2$. From Lemma 4.3.9, T is diagonalizable. For $\beta := ((1, 2), (2, -1))$, we have

$$[T]_\beta^\beta = \operatorname{diag}(1, -1) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Note that $[T^2]_\beta^\beta = I_2 = [I_{\mathbf{R}^2}]_\beta^\beta$, so $T^2 = I_{\mathbf{R}^2}$. The point of this example is that, once we can diagonalize T, taking powers of T becomes very easy.

Definition 4.3.13 (**Diagonalizable Matrix**)**.** An $n \times n$ matrix A is **diagonalizable** if the corresponding linear transformation L_A is diagonalizable.

Lemma 4.3.14. *A matrix A is diagonalizable if and only if there exists an invertible matrix Q and a diagonal matrix D such that $A = QDQ^{-1}$. That is, a matrix A is diagonalizable if and only if it is similar to a diagonal matrix.*

Proof. Suppose A is an $n \times n$ diagonalizable matrix. Let β denote the standard basis of $\mathbf{F}^n$, so that $A = [L_A]_\beta^\beta$. Since A is diagonalizable, there exists an ordered basis β' such that $D := [L_A]_{\beta'}^{\beta'}$ is diagonal. From Lemma 2.7.3, there exists an invertible matrix $Q := [I_{\mathbf{F}^n}]_{\beta'}^\beta$ such that

$$A = [L_A]_\beta^\beta = Q[L_A]_{\beta'}^{\beta'}Q^{-1} = QDQ^{-1}.$$

We now prove the converse. Suppose $A = QDQ^{-1}$, where Q is invertible and D is diagonal. Let $\lambda_1, \ldots, \lambda_n$ such that $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Then $De_i = \lambda_i e_i$ for all $i \in \{1, \ldots, n\}$, so

$$A(Qe_i) = QDQ^{-1}Qe_i = QDe_i = \lambda_i Qe_i.$$

So, Qe_i is an eigenvector of A, for each $i \in \{1, \ldots, n\}$. Since Q is invertible and $(e_1, \ldots, e_n)$ is a basis of $\mathbf{F}^n$, we see that $(Qe_1, \ldots, Qe_n)$ is also a basis of $\mathbf{F}^n$. So, $\beta'' := (Qe_1, \ldots, Qe_n)$ is a basis of $\mathbf{F}^n$ consisting of eigenvectors of A, so A is diagonalizable by Lemma 4.3.9, since $[L_A]_{\beta''}^{\beta''}$ is diagonal. □

Lemma 4.3.15. *Let A be an $n \times n$ matrix. Suppose $\beta' = (v_1, \ldots, v_n)$ is an ordered basis of $\mathbf{F}^n$ such that v_i is an eigenvector of A with eigenvalue λ_i for all $i \in \{1, \ldots, n\}$. Let Q be the matrix with columns $v_1, \ldots, v_n$ (where we write each v_i in the standard basis). Then*

$$A = Q \operatorname{diag}(\lambda_1, \ldots, \lambda_n)Q^{-1}.$$

Proof. Let β be the standard basis of $\mathbf{F}^n$. Note that $[I_{\mathbf{F}^n}]_{\beta'}^\beta = Q$. So, by Lemma 2.7.3,

$$A = [L_A]_\beta^\beta = Q[L_A]_{\beta'}^{\beta'}Q^{-1}.$$

Since $L_A v_i = \lambda_i v_i$ for all $i \in \{1, \ldots, n\}$,

$$[L_A]_{\beta'}^{\beta'} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n).$$

The Lemma follows. □

4.4. **Characteristic Polynomial.**

Lemma 4.4.1. *Let A be an $n \times n$ matrix. Then $\lambda \in \mathbf{F}$ is an eigenvalue of A if and only if* $\det(A - \lambda I_n) = 0$.

Proof. Suppose λ is an eigenvalue of A. Then there exists $v \in \mathbf{F}^n$ such that $Av = \lambda v$ and $v \neq 0$, so that $(A - \lambda I_n)v = 0$. So, $(A - \lambda I_n)$ is not invertible, and $\det(A - \lambda I_n) = 0$, from the contrapositive of Theorem 3.4.5. Conversely, if $\det(A - \lambda I_n) = 0$, then $A - \lambda I_n$ is not invertible, from the contrapositive of Theorem 3.4.5. In particular, $A - \lambda I_n$ is not one-to-one. So, there exists $v \in \mathbf{F}^n$ with $v \neq 0$ such that $(A - \lambda I_n)v = 0$, i.e. $Av = \lambda v$. □

Definition 4.4.2 (**Characteristic Polynomial**)**.** Let A be an $n \times n$ with entries in a field $\mathbf{F}$. Let $\lambda \in \mathbf{F}$, and define the **characteristic polynomial** $f(\lambda)$ of A, by

$$f(\lambda) := \det(A - \lambda I_n).$$

Lemma 4.4.3. *Let A, B be similar matrices. Then A, B have the same characteristic polynomial.*

Proof. Let $\lambda \in \mathbf{F}$. Since A, B are similar, there exists an invertible matrix Q such that $A = QBQ^{-1}$. So, using the multiplicative property of the determinant,

$$\begin{aligned}\det(A - \lambda I) &= \det(QBQ^{-1} - \lambda I) = \det(Q(B - \lambda I)Q^{-1}) \\ &= \det(Q)\det(B - \lambda I)\det(Q^{-1}) = \det(Q)\det(Q)^{-1}\det(B - \lambda I) \\ &= \det(B - \lambda I).\end{aligned}$$

□

Lemma 4.4.4. *Let A be an $n \times n$ matrix all of whose entries lie in $P_1(\mathbf{F})$. Then $\det(A) \in P_n(\mathbf{F})$.*

Proof. From Exercise 4.0.1 from the homework, $\det(A)$ is a sum of polynomials of degree at most n. That is, $\det(A)$ itself is in $P_n(\mathbf{R})$. □

Remark 4.4.5. From this Lemma, we see that the characteristic polynomial of A is a polynomial of degree at most n.

Lemma 4.4.6. *Let A be an $n \times n$ matrix with entries A_{ij}, $i, j \in \{1, \ldots, n\}$. Then there exists $g \in P_{n-2}(\mathbf{F})$ such that*

$$f(\lambda) = \det(A - \lambda I) = (A_{11} - \lambda) \cdots (A_{nn} - \lambda) + g(\lambda)$$

Proof. Let $B := A - \lambda I$. From Exercise 4.0.1 from the homework,

$$\det(A - \lambda I) = \prod_{i=1}^{n}(A_{ii} - \lambda) + \sum_{\sigma \in S_n \,:\, \sigma \neq I_n} \operatorname{sign}(\sigma) \prod_{i=1}^{n} B_{i\sigma(i)}.$$

Note that each term in the sum on the right has a number of λ terms equal to the number of $i \in \{1, \ldots, n\}$ such that $i = \sigma(i)$. So, if $\sigma \in S_n$ and $\sigma \neq I_n$, it suffices to show that there exist at least two integers $i, j \in \{1, \ldots, n\}$ with $i \neq j$ such that $\sigma(i) \neq i$ and $\sigma(j) \neq j$. We prove this assertion by contradiction. Suppose there exists $\sigma \in S_n$, $\sigma \neq I_n$ with exactly one $i \in \{1, \ldots, n\}$ with $\sigma(i) \neq i$. Then $\sigma(k) = k$ for all $k \in \{1, \ldots, n\} \setminus \{i\}$. Since σ is a permutation, σ is onto, so there exists $i' \in \{1, \ldots, n\}$ such that $\sigma(i') = i$. Since $\sigma(k) = k$ for all $k \in \{1, \ldots, n\} \setminus \{i\}$, we must therefore have $i = i'$, so that $\sigma(i) = i$, a contradiction. We conclude that since $\sigma \neq I_n$, there exist at least two $i, j \in \{1, \ldots, n\}$ with $i \neq j$ such that $\sigma(i) \neq i$ and $\sigma(j) \neq j$, as desired. □

Definition 4.4.7 (**Trace**)**.** Let A be an $n \times n$ matrix with entries A_{ij}, $i, j \in \{1, \ldots, n\}$. Then the **trace** of A, denoted by $\operatorname{Tr}(A)$, is defined as

$$\operatorname{Tr}(A) := \sum_{i=1}^{n} A_{ii}.$$

Theorem 4.4.8. *Let A be an $n \times n$ matrix. There exist scalars $a_1, \ldots, a_{n-2} \in \mathbf{F}$ such that the characteristic polynomial $f(\lambda)$ of A satisfies*

$$f(\lambda) = (-1)^n \lambda^n + (-1)^{n-1}\operatorname{Tr}(A)\lambda^{n-1} + a_{n-2}\lambda^{n-2} + \cdots + a_1\lambda + \det(A).$$

Proof. From Lemma 4.4.6, there exists $g \in P_{n-2}(\mathbf{F})$ such that

$$f(\lambda) = (A_{11} - \lambda) \cdots (A_{nn} - \lambda) + g(\lambda).$$

Multiplying out the product terms, we therefore get the two highest order terms of f. That is, there exists $G \in P_{n-2}(\mathbf{F})$ such that

$$f(\lambda) = (-\lambda)^n + \mathrm{Tr}(A)(-\lambda)^{n-1} + G(\lambda).$$

Finally, to get the zeroth order term of the polynomial f, note that by definition of the characteristic polynomial, $f(0) = \det(A)$. □

Example 4.4.9. Let $a, b, c, d \in \mathbf{R}$. Then the characteristic polynomial of

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is

$$(a-\lambda)(d-\lambda) - bc = \lambda^2 - \lambda(a+d) + (ad-bc) = \lambda^2 - \lambda\mathrm{Tr}(A) + \det(A).$$

Example 4.4.10. The characteristic polynomial of

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is $\lambda^2 - 1 = (\lambda+1)(\lambda-1)$.

Example 4.4.11. Let $i := \sqrt{-1}$. The characteristic polynomial of

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

is $\lambda^2 + 1 = (\lambda+i)(\lambda-i)$. However, we cannot factor λ^2+1 using only real numbers. So, as we will see below, we can diagonalize this matrix over the complex numbers, but not over the real numbers.

Theorem 4.4.12 (**The Fundamental Theorem of Algebra**). *Let $f(\lambda)$ be a real polynomial of degree n. Then there exist $\lambda_0, \lambda_1, \ldots \lambda_n \in \mathbf{C}$ such that*

$$f(\lambda) = \lambda_0 \prod_{i=1}^{n} (\lambda - \lambda_i).$$

Remark 4.4.13. This theorem is one of the reasons that complex numbers are useful. If we have complex numbers, then any real matrix has a characteristic polynomial that can be factored into complex roots. Without complex numbers, we could not do this.

4.5. **Diagonalizability.** Recall that n linearly independent vectors in $\mathbf{F}^n$ form a basis of $\mathbf{F}^n$. So, from Lemma 4.3.9 or the proof of Lemma 4.3.14, we have

Lemma 4.5.1. *Let A be an $n \times n$ matrix with elements in $\mathbf{F}$. Then A is diagonalizable (over $\mathbf{F}$) if and only if there exists a set $(v_1, \ldots, v_n)$ of linearly independent vectors in $\mathbf{F}^n$ such that v_i is an eigenvector of A for all $i \in \{1, \ldots, n\}$.*

We now examine some ways of finding a set of linearly independent eigenvectors of A, since this will allow us to diagonalize A.

Proposition 4.5.2. *Let A be an $n \times n$ matrix. Let $v_1, \ldots, v_k$ be eigenvectors of A with eigenvalues $\lambda_1, \ldots, \lambda_k$, respectively. If $\lambda_1, \ldots, \lambda_k$ are all distinct, then the vectors $v_1, \ldots, v_k$ are linearly independent.*

Proof. We argue by contradiction. Assume there exist $\alpha_1, \ldots, \alpha_k \in \mathbf{F}$ not all zero such that

$$\sum_{i=1}^{k} \alpha_i v_i = 0.$$

Without loss of generality, $\alpha_1 \neq 0$. Applying $(A - \lambda_k I)$ to both sides,

$$0 = \sum_{i=1}^{k-1} \alpha_i (A - \lambda_k I) v_i = \sum_{i=1}^{k-1} \alpha_i (\lambda_i - \lambda_k) v_i.$$

We now apply $(A - \lambda_{k-1} I)$ to both sides, and so on. Continuing in this way, we eventually get the equality

$$0 = \alpha_1 (\lambda_1 - \lambda_k)(\lambda_1 - \lambda_{k-1}) \cdots (\lambda_1 - \lambda_2) v_1.$$

Since $\lambda_1, \ldots, \lambda_k$ are all distinct, and $\alpha_1 \neq 0$, and since $v_1 \neq 0$ (since it is an eigenvector), we have arrived at a contradiction. We conclude that $v_1, \ldots, v_k$ are linearly independent. □

Corollary 4.5.3. *Let A be an $n \times n$ matrix with elements in $\mathbf{F}$. Suppose the characteristic polynomial $f(\lambda)$ of A can be written as $f(\lambda) = \prod_{i=1}^{n} (\lambda_i - \lambda)$, where $\lambda_i \in \mathbf{F}$ are distinct, for all $i \in \{1, \ldots, n\}$. Then A is diagonalizable.*

Proof. For all $i \in \{1, \ldots, n\}$, let $v_i \in \mathbf{F}^n$ be the eigenvector corresponding to the eigenvalue λ_i. Setting $k = n$ in Proposition 4.5.2 shows that $v_1, \ldots, v_n$ are linearly independent. Lemma 4.5.1 therefore completes the proof. □

Example 4.5.4. Consider

$$A = \begin{pmatrix} 1 & -2 \\ 1 & 4 \end{pmatrix}.$$

The characteristic polynomial is then

$$f(\lambda) = (1 - \lambda)(4 - \lambda) + 2 = \lambda^2 - 5\lambda + 6 = (\lambda - 2)(\lambda - 3).$$

So, $f(\lambda)$ has two distinct real roots, and we can diagonalize A over $\mathbf{R}$. Observe that $v_1 = (2, -1)$ is an eigenvector with eigenvalue 2 and $v_2 = (1, -1)$ is an eigenvector with eigenvalue 3. So, if we think of the eigenvectors as column vectors, and use them to define Q,

$$Q := \begin{pmatrix} 2 & 1 \\ -1 & -1 \end{pmatrix},$$

we then have the desired diagonalization

$$\begin{pmatrix} 1 & -2 \\ 1 & 4 \end{pmatrix} = Q \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} Q^{-1} = \begin{pmatrix} 2 & 1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & -2 \end{pmatrix}.$$

Exercise 4.5.5. Using the matrix from Example 4.4.11, find its diagonalization over $\mathbf{C}$.

In summary, if A is an $n \times n$ matrix with elements in $\mathbf{F}$, and if we can write the characteristic polynomial of A as a product of n distinct roots in $\mathbf{F}$, then A is diagonalizable over $\mathbf{F}$. On the other hand, if we cannot write the characteristic polynomial as a product of n roots in $\mathbf{F}$, then A is not diagonalizable over $\mathbf{F}$. (Combining Lemmas 4.3.14 and 4.4.3 shows that, if A is diagonalizable, then it has the same characteristic polynomial as a diagonal matrix. That is, the characteristic polynomial of A is the product of n roots.) (Recalling Example 4.4.11, the real matrix with characteristic polynomial $\lambda^2 + 1$ can be diagonalized over $\mathbf{C}$ but not over $\mathbf{R}$.)

The only remaining case to consider is when the characteristic polynomial of A can be written as a product of n non-distinct roots of $\mathbf{F}$. Unfortunately, this case is more complicated. It can be dealt with, but we don't have to time to cover the entire topic. The two relevant concepts here would be the Jordan normal form and the minimal polynomial.

To see the difficulty, note that the matrix

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

is diagonal, so it is diagonalizable. Also, the standard basis of $\mathbf{R}^2$ are eigenvectors, and the characteristic polynomial is $(2-\lambda)^2$.

On the other hand, consider the matrix

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

This matrix also has characteristic polynomial $(2-\lambda)^2$, but it is not diagonalizable. To see this, we will observe that the eigenvectors of A do not form a basis of $\mathbf{R}^2$. Since 2 is the only eigenvalue, all of the eigenvectors are in the null space of

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

However, this matrix has only a one-dimensional null space, which is spanned by the column vector $(1, 0)$. Since the eigenvectors of A do not form a basis of $\mathbf{R}^2$, A is not diagonalizable by Remark 4.3.11 (or Lemma 4.3.9).

5. Inner Products, Adjoints, Spectral Theorems, Self-Adjoint Operators

5.2. **Inner Product Spaces.** Up until this point, we have focused on the linear properties of vector spaces. Our investigation now becomes much deeper when we consider more geometric properties of vector spaces. That is we now start to consider the size of vectors, and how close one vector can be to another. These notions are made rigorous by the introduction of norms and inner products, respectively. Also, with this more geometric information, linear algebra and analysis will start to relate much more with each other.

We first introduce the concept of a general norm, which measures the length of vectors. We then introduce the more specific concept of an inner product, which measures the angle between two vectors.

Definition 5.2.1 (**Normed Linear Space**)**.** Let $\mathbf{F}$ denote either $\mathbf{R}$ or $\mathbf{C}$. Let V be a vector space over $\mathbf{F}$. A **normed linear space** is a vector space V equipped with a **norm**. A norm is a function $V \to \mathbf{R}$, denoted by $\|\cdot\|$, which satisfies the following properties.

(a) For all $v \in V$, for all $\alpha \in \mathbf{F}$, $\|\alpha v\| = |\alpha| \, \|v\|$. (Homogeneity)
(b) For all $v \in V$ with $v \neq 0$, $\|v\|$ is a positive real number; $\|v\| > 0$. And $v = 0$ if and only if $\|v\| = 0$. (Positive definiteness)
(c) For all $v, w \in V$, $\|v + w\| \leq \|v\| + \|w\|$. (Triangle Inequality)

Example 5.2.2. Let $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$. Define the 2-norm on $\mathbf{R}^n$ by

$$\|x\|_2 := \sqrt{x_1^2 + \cdots + x_n^2}.$$

So, for $n = 1$, we have $\|x\|_2 = |x|$. We will see below one way to show the triangle inequality for the 2-norm.

Define the 1-norm on $\mathbf{R}^n$ by

$$\|x\|_1 := \sum_{i=1}^{n} |x_i| .$$

Define the ∞-norm on $\mathbf{R}^n$ by

$$\|x\|_\infty := \max_{i=1,\ldots,n} |x_i| .$$

Exercise 5.2.3. Let $x, y \in \mathbf{R}$. Verify that $|x + y| \leq |x| + |y|$. Deduce that the triangle inequality holds for the 1-norm and the ∞-norm.

Exercise 5.2.4. Let V be a normed linear space. Show that, for any $v, w \in V$, $\|v - w\| \geq |\|v\| - \|w\||$.

Definition 5.2.5 (**Complex Conjugate**)**.** Let $i := \sqrt{-1}$. Let $x, y \in \mathbf{R}$, and let $z = x+iy \in \mathbf{C}$. Define $\overline{z} := x - iy$. Define $|z| := \sqrt{x^2 + y^2}$. Note that $|z|^2 = z\overline{z}$.

Definition 5.2.6 (**Inner Product**)**.** Let $\mathbf{F}$ denote either $\mathbf{R}$ or $\mathbf{C}$. Let V be a vector space over $\mathbf{F}$. An **inner product space** is a vector space V equipped with an **inner product**. An inner product is a function $V \times V \to \mathbf{F}$, denoted by $\langle \cdot, \cdot \rangle$, which satisfies the following properties.

(a) For all $v, v', w \in V$, $\langle v + v', w \rangle = \langle v, w \rangle + \langle v', w \rangle$. (Linearity in the first argument).
(b) For all $v, w \in V$, for all $\alpha \in \mathbf{F}$, $\langle \alpha v, w \rangle = \alpha \langle v, w \rangle$. (Homogeneity in the first argument)
(c) For all $v \in V$, if $v \neq 0$, then $\langle v, v \rangle$ is a positive real number; $\langle v, v \rangle > 0$. (Positivity)
(d) For all $v, w \in V$, $\langle v, w \rangle = \overline{\langle w, v \rangle}$. (Conjugate symmetry)

Exercise 5.2.7. Using the above properties, show the following things.

(e) For all $v, v', w \in V$, $\langle w, v + v' \rangle = \langle w, v \rangle + \langle w, v' \rangle$. (Linearity in the second argument)
(f) For all $v, w \in V$, for all $\alpha \in \mathbf{F}$, $\langle v, \alpha w \rangle = \overline{\alpha} \langle v, w \rangle$.
(g) For all $v \in V$, $\langle v, 0 \rangle = \langle 0, v \rangle = 0$.
(h) $\langle v, v \rangle = 0$ if and only if $v = 0$.

Remark 5.2.8. If $\mathbf{F} = \mathbf{R}$, then property (d) says that $\langle v, w \rangle = \langle w, v \rangle$.

Example 5.2.9. Let $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$, and let $y = (y_1, \ldots, y_n) \in \mathbf{R}^n$. Define the standard inner product (or dot product) on $\mathbf{R}^n$ by

$$\langle x, y \rangle := \sum_{i=1}^{n} x_i y_i.$$

More generally, if $\alpha_1, \ldots, \alpha_n > 0$, then the following definition also gives an inner product

$$\langle x, y \rangle_\alpha := \sum_{i=1}^{n} \alpha_i x_i y_i.$$

Example 5.2.10. Let $w = (w_1, \ldots, w_n) \in \mathbf{C}^n$, and let $z = (z_1, \ldots, z_n) \in \mathbf{C}^n$. Define the standard inner product (or dot product) on $\mathbf{C}^n$ by

$$\langle w, z \rangle := \sum_{i=1}^{n} w_i \overline{z_i}.$$

Example 5.2.11. Let $A, B \in M_{n\times n}(\mathbf{R})$. Define the standard inner product on $M_{n\times n}(\mathbf{R})$ by

$$\langle A, B \rangle := \mathrm{Tr}(B^t A).$$

Example 5.2.12. Let $f, g \in C([0,1], \mathbf{R})$. That is, f, g are continuous real valued functions on $[0,1]$. Define

$$\langle f, g \rangle := \int_0^1 f(t)g(t)dt.$$

Definition 5.2.13 (**Orthogonal Vectors**)**.** Let V be an inner product space, and let $v, w \in V$. We say that v, w are **orthogonal** if $\langle v, w \rangle = 0$.

Lemma 5.2.14 (**The Cauchy-Schwarz Inequality**)**.** *Let V be an inner product space. Then, for any $v, w \in V$,*

$$|\langle v, w \rangle| \leq \sqrt{\langle v, v \rangle}\sqrt{\langle w, w \rangle}.$$

Proof. If $w = 0$, then both sides of our inequality are zero, so the inequality holds, and we may assume $w \neq 0$. So, $\langle w, w \rangle > 0$. Define $\alpha := \langle v, w \rangle / \langle w, w \rangle$. Note that by conjugate-symmetry of the inner product,

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

Also, by positivity of the inner product, $\langle v - \alpha w, v - \alpha w \rangle \geq 0$. That is,

$$\langle v, v \rangle - \alpha \langle w, v \rangle - \overline{\alpha} \langle v, w \rangle + |\alpha|^2 \langle w, w \rangle \geq 0$$

Substituting in the definition of α,

$$\langle v, v \rangle - |\langle v, w \rangle|^2 / \langle w, w \rangle - |\langle v, w \rangle|^2 / \langle w, w \rangle + |\langle v, w \rangle|^2 / \langle w, w \rangle \geq 0.$$

Simplifying, we get $\langle v, v \rangle - |\langle v, w \rangle|^2 / \langle w, w \rangle \geq 0$, as desired. □

Remark 5.2.15. If the choice of α looks a bit mysterious, note that if $v = \beta w$ for some $\beta \in \mathbf{R}$, then the Cauchy-Schwarz inequality becomes an equality. And if $\langle v, w \rangle = 0$, then the Cauchy-Schwarz inequality has zero on the left side and a possibly large number on the right. So, the positive number $\sqrt{\langle v, v \rangle}\sqrt{\langle w, w \rangle} - |\langle v, w \rangle|$ somehow measures how close v and w are to being parallel. Also, in the proof of the Cauchy-Schwarz inequality, αw is parallel to w and $v - \alpha w$ is orthogonal to w, so writing

$$v = (v - \alpha w) + \alpha w,$$

we see that the size of $v - \alpha w$ also measures how close v and w are to being parallel.

We now show that an inner product space is a normed linear space, with norm $\|v\| := \sqrt{\langle v, v \rangle}$.

Lemma 5.2.16. *Let $\langle , \rangle$ be an inner product on a vector space V. Then the function $\|\cdot\| : V \to \mathbf{R}$ defined by $\|v\| := \sqrt{\langle v, v \rangle}$ is a norm on V.*

Proof. Homogeneity and positive definiteness follow readily from the definition of the inner product, and from Exercise 5.2.7(h). It therefore suffices to show that the triangle inequality holds. Let $v, w \in V$. We need to show that $\|v+w\| \leq \|v\| + \|w\|$. In order to show this inequality, it suffices to show that its square holds.

$$\begin{aligned}
\|v+w\|^2 &= |\langle v+w, v+w\rangle| = |\langle v, v\rangle + \langle w, w\rangle + \langle v, w\rangle + \langle w, v\rangle| \\
&\leq |\langle v, v\rangle| + |\langle w, w\rangle| + |\langle v, w\rangle| + |\langle w, v\rangle| \qquad \text{, by Exercise 5.2.3} \\
&\leq \|v\|^2 + \|w\|^2 + 2\|v\|\,\|w\| \qquad \text{, by Lemma 5.2.14} \\
&= (\|v\| + \|w\|)^2.
\end{aligned}$$

□

Consequently, the triangle inequality holds for the norm $\|\cdot\|_2$ on $\mathbf{R}^n$, for any $n \geq 1$.

5.3. **Orthogonality.** Let V be an inner product space with inner product $\langle\,,\rangle$. Recall that $v, w \in V$ are said to be orthogonal if $\langle v, w\rangle = 0$. If v, w are orthogonal, we also sometimes say that v, w are perpendicular, and we write $v \perp w$.

Lemma 5.3.1. *Let V be an inner product space. Suppose $v \in V$ is orthogonal to each of the vectors $v_1, \ldots, v_n \in V$. Then v is orthogonal to any linear combination of $v_1, \ldots, v_n$.*

Proof. Since $\langle v, v_i\rangle = 0$ for all $i \in \{1, \ldots, n\}$, if $\alpha_1, \ldots, \alpha_n \in \mathbf{F}$, we have

$$\langle v, \sum_{i=1}^{n} \alpha_i v_i\rangle = \sum_{i=1}^{n} \overline{\alpha_i}\langle v, v_i\rangle = 0.$$

□

Theorem 5.3.2 (**Pythagorean Theorem**). *Let V be an inner product space, and let $v, w \in V$ be orthogonal. Then $\|v+w\|^2 = \|v\|^2 + \|w\|^2$.*

Proof.

$$\|v+w\|^2 = \langle v+w, v+w\rangle = \langle v, v\rangle + \langle w, w\rangle + \langle v, w\rangle + \langle w, v\rangle = \|v\|^2 + \|w\|^2.$$

□

Theorem 5.3.3 (**Generalized Pythagorean Theorem**). *Let V be an inner product space, and let $v_1, \ldots, v_n \in V$ be orthogonal to each other. That is, $\langle v_i, v_j\rangle = 0$ for all $i, j \in \{1, \ldots, n\}$ with $i \neq j$. Then*

$$\left\|\sum_{i=1}^{n} v_i\right\|^2 = \sum_{i=1}^{n} \|v_i\|^2$$

Proof. We induct on n. The base case has been proven, so we only need to prove the inductive step. Suppose the assertion is true for a fixed n. Then, let $v_1, \ldots, v_{n+1}$ be orthogonal to each other. From Lemma 5.3.1, v_{n+1} is orthogonal to $\sum_{i=1}^{n} v_i$. So, applying the Pythagorean Theorem to v_{n+1} and $\sum_{i=1}^{n} v_i$, and then using the inductive hypothesis,

$$\left\|v_{n+1} + \sum_{i=1}^{n} v_i\right\|^2 = \|v_{n+1}\|^2 + \left\|\sum_{i=1}^{n} v_i\right\|^2 = \|v_{n+1}\|^2 + \sum_{i=1}^{n} \|v_i\|^2.$$

□

Corollary 5.3.4. *Let V be an inner product space, and let $v_1, \ldots, v_n \in V$ be orthogonal to each other. That is, $\langle v_i, v_j \rangle = 0$ for all $i, j \in \{1, \ldots, n\}$ with $i \neq j$. Let $\alpha_1, \ldots, \alpha_n \in \mathbf{F}$. Then*

$$\left\| \sum_{i=1}^{n} \alpha_i v_i \right\|^2 = \sum_{i=1}^{n} |\alpha_i|^2 \|v_i\|^2 .$$

Proof. If $\langle v_i, v_j \rangle = 0$, then $\langle \alpha_i v_i, \alpha_j v_j \rangle = 0$. So, apply Theorem 5.3.3 to the set of vectors $\alpha_1 v_1, \ldots, \alpha_n v_n$. □

Definition 5.3.5 (**Orthogonal Set, Orthonormal Set**)**.** Let V be an inner product space and let $(v_1, \ldots, v_n)$ be a collection of vectors in V. The set of vectors $(v_1, \ldots, v_n)$ is said to be **orthogonal** if $\langle v_i, v_j \rangle = 0$ for all $i, j \in \{1, \ldots, n\}$ with $i \neq j$. If additionally $\langle v_i, v_i \rangle = 1$ for all $i \in \{1, \ldots, n\}$, the set of vectors $(v_1, \ldots, v_n)$ is called **orthonormal**.

Corollary 5.3.6. *Let V be an inner product space, and let $v_1, \ldots, v_n \in V$ be an orthonormal set of vectors. Then*

$$\left\| \sum_{i=1}^{n} \alpha_i v_i \right\|^2 = \sum_{i=1}^{n} |\alpha_i|^2 .$$

Corollary 5.3.7. *Any set of orthonormal vectors is linearly independent.*

5.3.1. *Orthonormal Bases.*

Definition 5.3.8 (**Orthonormal Basis**)**.** Let V be an inner product space. An **orthonormal basis** of V is a collection $(v_1, \ldots, v_n)$ of orthonormal vectors that is also a basis for V.

Corollary 5.3.9. *Let V be an n-dimensional inner product space. Let $(v_1, \ldots, v_n)$ be an orthonormal set in V. Then $(v_1, \ldots, v_n)$ is an orthonormal basis of V.*

Proof. By Corollary 5.3.7, $(v_1, \ldots, v_n)$ is linearly independent. By Corollary 1.6.14(e), If we have n linearly independent vectors in an n-dimensional space, then these vectors form a basis of V. □

Theorem 5.3.10. *Let V be an inner product space. Let $(v_1, \ldots, v_n)$ be an orthonormal basis of V. Then, for any $v \in V$, we have*

$$v = \sum_{i=1}^{n} \langle v, v_i \rangle v_i.$$

Proof. Let $v \in V$. Since $(v_1, \ldots, v_n)$ is a basis of V, there exist $\alpha_1, \ldots, \alpha_n \in \mathbf{F}$ such that

$$v = \sum_{i=1}^{n} \alpha_i v_i. \qquad (*)$$

So, we need to show that $\alpha_i = \langle v, v_i \rangle$ for all $i \in \{1, \ldots, n\}$. Let $j \in \{1, \ldots, n\}$. Taking the inner product of both sides of $(*)$ with v_j, and then applying orthonormality,

$$\langle v, v_j \rangle = \left\langle \sum_{i=1}^{n} \alpha_i v_i, v_j \right\rangle = \sum_{i=1}^{n} \alpha_i \langle v_i, v_j \rangle = \alpha_i \langle v_j, v_j \rangle = \alpha_j.$$

□

Corollary 5.3.11. *Let V be an inner product space. Let $\beta = (v_1, \ldots, v_n)$ be an orthonormal basis of V. Then, the coordinate vector $[v]^\beta$ has the form*

$$[v]^\beta = \begin{pmatrix} \langle v, v_1 \rangle \\ \vdots \\ \langle v, v_n \rangle \end{pmatrix}.$$

Remark 5.3.12. Let V, W be finite-dimensional inner product spaces. Let $\beta = (v_1, \ldots, v_n)$ be an orthonormal basis of V and let γ be an orthonormal basis of W. Let $T\colon V \to W$ be a linear transformation. Then we can compute $[T]^\gamma_\beta$ using inner products, since its columns are $[T(v_1)]^\gamma, \ldots, [T(v_n)]^\gamma$.

Example 5.3.13. Consider $C([0,1], \mathbf{C})$. As usual, let $i := \sqrt{-1}$. Let $f, g \in C([0,1], \mathbf{C})$, and consider the standard inner product

$$\langle f, g \rangle := \int_0^1 f(t)\overline{g(t)}dt.$$

Let $k \in \mathbf{Z}$, $t \in [0,1]$. Define $v_k(t) := e^{2\pi ikt}$. We claim that the set $\{v_k\}_{k\in\mathbf{Z}}$ is an orthonormal set. (In a suitable sense, it is also an orthonormal basis, but we cannot cover this topic here; for more, look into Fourier analysis.) Let $j, k \in \mathbf{Z}$ and observe

$$\langle v_j, v_k \rangle = \int_0^1 v_j(t)\overline{v_k(t)}dt = \int_0^1 e^{2\pi ijt}e^{-2\pi ikt}dt = \int_0^1 e^{2\pi i(j-k)t}dt$$

So, if $j = k$, we get $\langle v_j, v_k \rangle = \int_0^1 dt = 1$. And if $j \neq k$, we have $j - k \in \mathbf{Z}$, so

$$\langle v_j, v_k \rangle = \frac{1}{2\pi i(j-k)}(e^{2\pi i(j-k)} - 1) = \frac{1}{2\pi i(j-k)}(1-1) = 0.$$

Let T_n denote the set of trigonometric polynomials of the form $a_0 + a_1e^{2\pi it} + \cdots + a_ne^{2\pi int}$, $a_1, \ldots, a_n \in \mathbf{C}$. Then T_n is a subspace of $C([0,1], \mathbf{C})$, so T_n is also an inner product space. So from Theorem 5.3.10, if $f \in T_n$, we have

$$f(t) = \sum_{j=0}^{n}(\int_0^1 f(s)e^{-2\pi ijs}ds)e^{2\pi ijt},$$

and from Corollary 5.3.6, we have **Plancherel's formula**

$$\int_0^1 |f(t)|^2 \, dt = \sum_{j=0}^{n} \left| \int_0^1 f(s)e^{-2\pi ijs}ds \right|^2.$$

The scalars $\int_0^1 f(s)e^{-2\pi ijs}ds$, $j \in \{0, \ldots, n\}$ are called the **Fourier coefficients** of f.

5.4. **Gram-Schmidt Orthogonalization.**

Definition 5.4.1 (**Unit Vector**)**.** Let V be a normed linear space, and let $v \in V$. If $\|v\| = 1$, we say that v is a **unit vector**.

Remark 5.4.2. Let $v \neq 0$. Then $v/\|v\|$ is a unit vector.

Definition 5.4.3 (**Projection onto a vector**)**.** Let v, w be vectors in an inner product space, with $w \neq 0$. Define the **orthogonal projection** of v onto w by

$$\mathrm{P}_w(v) := \frac{\langle v, w\rangle}{\langle w, w\rangle} w = \left\langle v, \frac{w}{\|w\|} \right\rangle \frac{w}{\|w\|}.$$

Note that P_w is a linear transformation.

As we saw in the proof of the Cauchy-Schwarz inequality, if $v, w \in V$ and $w \neq 0$, we can write

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

And $v - P_w(v)$ is orthogonal to w, while $P_w(v)$ is parallel to w. Therefore, $v - P_w(v)$ is orthogonal to $P_w(v)$.

Definition 5.4.4 (**Projection onto a subspace**)**.** Let V be an inner product space. Let $W \subseteq V$ be an n-dimensional subspace of V. Let $w_1, \ldots, w_n$ be an orthogonal set of nonzero vectors in W. Let $v \in V$. Define the **orthogonal projection** of v onto W by

$$P_W(v) := \sum_{i=1}^{n} \left\langle v, \frac{w_i}{\|w_i\|} \right\rangle \frac{w_i}{\|w_i\|}.$$

Note that $P_W \colon V \to V$ is a linear transformation, and $R(P_W) \subseteq W$.

Remark 5.4.5. $P_W(v) = v$ if and only if $v \in W$ by Theorem 5.3.10. Also, the definition of $P_W(v)$ does not depend on the orthogonal set of nonzero vectors $w_1, \ldots, w_n$. This follows by applying Theorem 5.3.10 to the orthonormal set $(w_1/\|w_1\|, \ldots, w_n/\|w_n\|)$.

Remark 5.4.6. Let $w_1, \ldots, w_n$ be an orthogonal set of nonzero vectors in W. As before, given $v \in V$ and W an n-dimensional subspace of V, we can write

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

Note that $P_W(v) \in W$, and $(v - P_W(v))$ is orthogonal to w_i for each $i \in \{1, \ldots, n\}$. So, by Lemma 5.3.1, $(v - P_W(v))$ is orthogonal to any vector in W.

Given a set of linearly independent vectors, we can create an orthonormal set of vectors from the linearly independent set by using projections and Remark 5.4.6. The procedure for creating these orthonormal sets is known as Gram-Schmidt orthogonalization.

Theorem 5.4.7 (**Gram-Schmidt Orthogonalization**)**.** *Let $v_1, \ldots, v_n$ be a linearly independent set of vectors in an inner product space V. Then we can create an orthogonal set of vectors in V as follows. Define*

$$w_1 := v_1.$$
$$w_2 := v_2 - P_{w_1}(v_2).$$
$$w_3 := v_3 - P_{\mathrm{span}(w_1, w_2)}(v_3).$$

And so on. In general, for $k \in \{2, \ldots, n\}$, define

$$w_k := v_k - P_{\mathrm{span}(w_1, \ldots, w_{k-1})}(v_k).$$

Then for each $k \in \{1, \ldots, n\}$, $(w_1, \ldots, w_k)$ is an orthogonal set of nonzero vectors in V. Also, $\mathrm{span}(w_1, \ldots, w_k) = \mathrm{span}(v_1, \ldots, v_k)$ for each $k \in \{1, \ldots, n\}$. Finally, note that the set $(w_1/\|w_1\|, \ldots, w_n/\|w_n\|)$ is an orthonormal set of vectors in V with the same span as $v_1, \ldots, v_n$.

Proof. Note that $w_2 \perp w_1$ from Remark 5.4.6. We will show that $\{w_1, \ldots, w_k\}$ is an orthogonal set of nonzero vectors, and $\text{span}(w_1, \ldots, w_k) = \text{span}(v_1, \ldots, v_k)$ by induction on k. Assume $\{w_1, \ldots, w_k\}$ is an orthogonal set of nonzero vectors, and $\text{span}(w_1, \ldots, w_k) = \text{span}(v_1, \ldots, v_k)$ for some k. Consider w_{k+1}. Using the definition of w_{k+1}, the inductive hypothesis, and Remark 5.4.5,

$$w_{k+1} = v_{k+1} - P_{\text{span}(w_1,\ldots,w_k)}(v_{k+1}) = v_{k+1} - P_{\text{span}(v_1,\ldots,v_k)}(v_{k+1}). \qquad (*)$$

By Remark 5.4.6, w_{k+1} is orthogonal to any vector in $\text{span}(v_1, \ldots, v_k) = \text{span}(w_1, \ldots, w_k)$. Also, $w_{k+1} \neq 0$, since $v_{k+1} \notin \text{span}(v_1, \ldots, v_k)$, by linear independence. That is, $v_{k+1} \neq P_{\text{span}(v_1,\ldots,v_k)}(v_{k+1})$ by Remark 5.4.5. Therefore, $\{w_1, \ldots, w_{k+1}\}$ is an orthogonal set of nonzero vectors. We now show the spanning property. From $(*)$ and the definition of the projection, $w_{k+1} \in \text{span}(v_1, \ldots, v_{k+1})$. So,

$$\text{span}(w_1, \ldots, w_{k+1}) \subseteq \text{span}(v_1, \ldots, v_{k+1}).$$

Now, note that the span on the right is $(k+1)$-dimensional by Corollary 5.3.7, as is the span on the left. So we must have equality. The induction step is complete, and we are done. $\square$

Example 5.4.8. Consider $P_2([-1, 1])$, the set of real polynomials of degree at most 2 on the interval $[-1, 1]$. Let $f, g \in P_2([-1, 1])$. We use the inner product

$$\langle f, g \rangle := \int_{-1}^{1} f(t)g(t)dt.$$

Let's start with the standard basis $(1, t, t^2)$ where t is a real variable, and let's create an orthonormal basis from the standard one. Define $v_1 := 1$, $v_2 := t$, and $v_3 := t^2$. Define

$$w_1 := v_1 = 1.$$

Note that $\langle w_1, w_1 \rangle = 2$, so w_1 has norm $\sqrt{2}$. Then, define

$$w_2 := v_2 - \langle v_2, w_1 \rangle w_1/2 = t.$$

We can verify that $\langle t, 1 \rangle = 0$. Note also that $\langle t, t \rangle = \int_{-1}^{1} t^2 dt = 2/3$, so $(\sqrt{3/2})w_2$ has norm 1. So, define

$$\begin{aligned} w_3 &:= v_3 - \langle v_3, w_2 \rangle w_2 / \|w_2\|^2 - \langle v_3, w_1 \rangle w_1 / \|w_1\|^2 \\ &= t^2 - (\int_{-1}^{1} x^3 dx)t(3/2) - (1/2)(\int_{-1}^{1} x^2 dx) \\ &= t^2 - 1/3. \end{aligned}$$

We can verify that $\langle t^2 - 1/3, 1 \rangle = 0$ and $\langle t^2 - 1/3, t \rangle = 0$. Also, $\int_{-1}^{1} (t^2 - 1/3)^2 dt = 8/45$, so $t^2 - 1/3$ has norm $\sqrt{8/45}$.

In conclusion, an orthonormal basis for $P_2([-1, 1])$, is

$$\left\{ \sqrt{\frac{1}{2}}, \sqrt{\frac{3}{2}}t, \sqrt{\frac{45}{8}}\left(t^2 - \frac{1}{3}\right) \right\}.$$

Corollary 5.4.9. *Every finite dimensional inner product space has an orthonormal basis.*

Proof. Recall from Definition 1.6.15 that every finite-dimensional vector space has a basis. Given this basis $v_1, \ldots, v_n$, apply the Gram-Schmidt Orthogonalization (Theorem 5.4.7) to get an orthonormal set $w_1/\|w_1\|, \ldots, w_n/\|w_n\|$. By Corollary 5.3.9, the vectors produced from the Gram-Schmidt process are an orthonormal basis. □

Corollary 5.4.10. *Let V be an inner product space, and let $W \subseteq V$ be a finite-dimensional subspace. Then there exists a linear transformation $P\colon V \to V$ such that $P^2 = P$, $R(P) \subseteq W$, and $P(w) = w$ for any $w \in W$. That is, P is a projection onto W.*

Proof. From Corollary 5.4.9, let $w_1, \ldots, w_n$ be an orthonormal basis for W. As in Definition 5.4.4, define

$$P(v) = P_W(v) := \sum_{i=1}^{n} \langle v, w_i \rangle w_i.$$

□

5.4.1. *Orthogonal Complements.*

Definition 5.4.11 (**Orthogonal Subspaces**)**.** Let V_1, V_2 be two subspaces of an inner product space V. If $v_1 \perp v_2$ for all $v_1 \in V_1, v_2 \in V_2$, we say that V_1 is **orthogonal** to V_2, and we write $V_1 \perp V_2$.

Lemma 5.4.12. *Let V_1, V_2 be two subspaces of an inner product space V. If $V_1 \perp V_2$, then $V_1 \cap V_2 = \{0\}$.*

Proof. Since V_1, V_2 are subspaces, $0 \in V_1$ and $0 \in V_2$, so $0 \in V_1 \cap V_2$. Now, let $v \in V_1 \cap V_2$. We will show that $v = 0$. Then $v \in V_2$. But since $v \in V_2$ and $V_2 \perp V_1$, we have $\langle v, v_1 \rangle = 0$ for all $v_1 \in V_1$. In particular, since $v \in V_1$, we have $\langle v, v \rangle = 0$. By the positive definiteness property of the inner product, we conclude that $v = 0$. That is, $V_1 \cap V_2 = \{0\}$. □

Definition 5.4.13 (**Orthogonal Complement**)**.** Let V_1 be a subspace of an inner product space V. Define the **orthogonal complement** of V_1 in V by

$$V_1^{\perp} := \{v \in V\colon \langle v, v_1 \rangle = 0, \forall\, v_1 \in V_1\}.$$

Exercise 5.4.14. Show that $\{0\}^{\perp} = V$ and $V^{\perp} = \{0\}$.

Exercise 5.4.15. Let V_1 be a subspace of an inner product space V. Show that $V_1^{\perp}$ is a subspace of V.

The following Theorem gives an algorithm for computing orthogonal complements.

Theorem 5.4.16. *Let V be an n-dimensional inner product space, and let $W \subseteq V$ be a k-dimensional subspace. Let $v_1, \ldots, v_k$ be a basis of W, and let $v_1, \ldots, v_n$ be an extension of that basis to V. (We proved that this extension exists in Corollary 1.6.14(f)). Let $w_1, \ldots, w_n$ be the orthonormal vectors produced by Gram-Schmidt orthogonalization. Then $w_1, \ldots, w_k$ is an orthonormal basis of W, and $w_{k+1}, \ldots, w_n$ is an orthonormal basis of $W^{\perp}$.*

Proof. From Theorem 5.4.7, $\operatorname{span}(w_1, \ldots, w_k) = \operatorname{span}(v_1, \ldots, v_k)$. Since W is k-dimensional, we conclude that $w_1, \ldots, w_k$ is a basis for W. Since $w_1, \ldots, w_k$ is also orthonormal, it is therefore an orthonormal basis of W.

Also, the vectors $w_{k+1}, \ldots, w_n$ are orthonormal, and therefore they are linearly independent by Corollary 5.3.7. So, it remains to show that $w_{k+1}, \ldots, w_n$ spans $W^{\perp}$. Let $j \in \{k +$

$1, \ldots, n\}$. By the Gram-Schmidt process, w_j is orthogonal to each of the vectors $w_1, \ldots, w_k$. By Lemma 5.3.1, w_j is then orthogonal to all of W. So, $w_j \in W^\perp$. So, $\text{span}(w_{k+1}, \ldots, w_n) \subseteq W^\perp$. It remains to show that every $w \in W^\perp$ is in $\text{span}(w_{k+1}, \ldots, w_n)$.

Let $w \in W^\perp$. Since $w \in V$ and $(w_1, \ldots, w_n)$ is an orthonormal basis of V, we have by Theorem 5.3.10,

$$w = \sum_{i=1}^{n} \langle w, w_i \rangle w_i.$$

Since $w \in W^\perp$, $\langle w, w_i \rangle = 0$ for each $i \in \{1, \ldots, k\}$. That is,

$$w = \sum_{i=k+1}^{n} \langle w, w_i \rangle w_i.$$

So, $w \in \text{span}(w_{k+1}, \ldots, w_n)$, as desired. □

Example 5.4.17. We continue the definitions and notation from Example 5.4.8. Consider $W \subseteq P_2([-1, 1])$, where W is the span of 1 and t. Let's compute $W^\perp$. To do this, we complete the set $(1, t)$ to a basis $(1, t, t^2)$. From Example 5.4.8, we then used Gram-Schmidt orthogonalization using $v_1 = 1$, $v_2 = t$ and $v_3 = t^2$. We found that the resulting orthonormal basis for $P_2([-1, 1])$ is

$$\left\{ \sqrt{\frac{1}{2}}, \sqrt{\frac{3}{2}} t, \sqrt{\frac{45}{8}} \left(t^2 - \frac{1}{3} \right) \right\}.$$

So, $\sqrt{1/2}$ and $\sqrt{(3/2)}t$ are an orthonormal basis of W. And therefore, $W^\perp$ is a one-dimensional space described as

$$W^\perp = \{\alpha(t^2 - 1/3) \colon \alpha \in \mathbf{R}\}.$$

Corollary 5.4.18 (Dimension Theorem for orthogonal complements). *Let W be a subspace of a finite-dimensional inner product space V. Then*

$$\dim(W) + \dim(W^\perp) = \dim(V).$$

Corollary 5.4.19. *Let W be a subspace of a finite-dimensional inner product space V. Then every $v \in V$ can be written uniquely as $v = w + n$ where $w \in W$ and $n \in W^\perp$.*

Proof. By Theorem 5.4.16, $\exists$ an orthonormal basis $w_1, \ldots, w_m$ of V such that $w_1, \ldots, w_k$ is an orthonormal basis of W, and $w_{k+1}, \ldots, w_m$ is an orthonormal basis of $W^\perp$. Let $v \in V$. by Theorem 5.3.10,

$$v = \sum_{i=1}^{m} \langle v, w_i \rangle w_i = \sum_{i=1}^{k} \langle v, w_i \rangle w_i + \sum_{i=k+1}^{m} \langle v, w_i \rangle w_i.$$

Define

$$w := \sum_{i=1}^{k} \langle v, w_i \rangle w_i, \qquad n := \sum_{i=k+1}^{m} \langle v, w_i \rangle w_i.$$

Then $v = w + n$, $w \in W$, $n \in W^\perp$. (As an aside, note that $w = P_W(v)$, and $n = v - P_W(v)$.) We now show the desired uniqueness statement. Suppose $v = w' + n'$ with $w' \in W$ and

$n' \in W^\perp$. We will be done once we show that $w = w'$ and $n = n'$. Since $w + n = w' + n'$, we have

$$w - w' = n - n'. \qquad (*)$$

The vector on the left of $(*)$ is in W, and the vector on the right of $(*)$ is in $W^\perp$. By Lemma 5.4.12, $W \cap W^\perp = \{0\}$. So, both sides of $(*)$ must be zero. That is, $w = w'$ and $n = n'$, as desired. □

Theorem 5.4.20 (**Orthogonal projections minimize length**). *Let W be a subspace of a finite-dimensional inner product space V. Let $v \in V$, and let $w = P_W(v)$ be the orthogonal projection of v onto W. Then, for any $w' \in W$ with $w' \neq w$, we have $\|v - w\| < \|v - w'\|$.*

Proof. From Corollary 5.4.19, write $v = w + n$, where $w := P_W(v) \in W$, and $n := v - w \in W^\perp$. Then $\|v - w\| = \|n\|$. Now, write

$$v - w' = (v - w) + (w - w').$$

Since $w, w' \in W$, $w - w' \in W$. Since $v - w = n \in W^\perp$, $\langle v - w, w - w' \rangle = 0$. So, by the Pythagorean Theorem (Theorem 5.3.2),

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

So, $\|v - w'\|^2 > \|v - w\|^2$ since $w \neq w'$, as desired. □

5.5. **Adjoints.**

5.5.1. *Linear Functionals.*

Definition 5.5.1 (**Linear Functional**). Let V be a vector space over a field $\mathbf{F}$. A **linear functional** is a linear transformation $T\colon V \to \mathbf{F}$.

Linear functionals are also known as dual vectors, covectors, or 1-forms. In order to understand some vector space V, it is often of interest to understand the set of all linear functionals on V. Such a classification becomes quite subtle especially for infinite dimensional spaces. However, as we show below, the case of finite-dimensional inner product spaces is fairly tame.

Example 5.5.2. Define $T\colon \mathbf{R}^3 \to \mathbf{R}$ by $T(a, b, c) := a + 2b + 3c$. Then T is a linear functional.

Example 5.5.3. Define $T\colon C([0, 1], \mathbf{R}) \to \mathbf{R}$ by $T(f) := \int_0^1 f(t)dt$. Then T is a linear functional.

Example 5.5.4. Define $T\colon C([0, 1], \mathbf{R}) \to \mathbf{R}$ by $T(f) := f(1/3)$. Then T is a linear functional.

Example 5.5.5. Let V be a finite-dimensional inner product space over $\mathbf{R}$. Let $w \in V$, and define $T\colon V \to \mathbf{R}$ by $T(v) := \langle v, w \rangle$. Then T is a linear functional.

As we now show, the previous example essentially classifies all linear functionals on a finite-dimensional inner product space.

Theorem 5.5.6 (**Riesz Representation Theorem**). *Let V be a finite-dimensional inner product space over a field $\mathbf{F}$. Let $T\colon V \to \mathbf{F}$ be a linear functional. Then there exists a unique vector $w \in V$ such that, for all $v \in V$, $T(v) = \langle v, w \rangle$.*

Proof. From Corollary 5.4.9, V has an orthonormal basis $v_1, \ldots, v_n$. Define

$$w := \sum_{j=1}^{n} v_j \overline{T(v_j)}.$$

Let $v \in V$. From Theorem 5.3.10,

$$v = \sum_{i=1}^{n} \langle v, v_i \rangle v_i.$$

Since T is linear, we get

$$T(v) = \sum_{i=1}^{n} \langle v, v_i \rangle T(v_i)$$

Since $v_1, \ldots, v_n$ is an orthonormal basis,

$$\langle v, w \rangle = \left\langle \sum_{i=1}^{n} \langle v, v_i \rangle v_i, \sum_{j=1}^{n} v_j \overline{T(v_j)} \right\rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} \langle v, v_i \rangle T(v_j) \langle v_i, v_j \rangle = \sum_{i=1}^{n} \langle v, v_i \rangle T(v_i) = T(v).$$

This completes the existence part of the proof. We now prove uniqueness.

Suppose there exists $w' \in V$ such that $T(v) = \langle v, w' \rangle$. We will show that $w = w'$. For all $v \in V$, $T(v) = \langle v, w \rangle = \langle v, w' \rangle$. That is,

$$\forall\, v \in V \ \langle v, w - w' \rangle = 0. \qquad (*)$$

Choosing $v = w - w'$ shows that $\langle w - w', w - w' \rangle = 0 = \|w - w'\|^2$. Therefore, $w - w' = 0$, as desired. □

5.5.2. *Adjoints.* Let $\mathbf{F}$ denote $\mathbf{R}$ or $\mathbf{C}$. Let V be a finite-dimensional inner product space over $\mathbf{F}$ with inner product $\langle\, , \rangle_V$, and let W be a finite-dimensional inner product space over $\mathbf{F}$ with inner product $\langle\, , \rangle_W$. Let $T \colon V \to W$ be a linear transformation. Given any $w \in W$, define a linear functional $T_w \colon V \to \mathbf{F}$ by

$$T_w(v) := \langle T(v), w \rangle_W.$$

Note that T_w is actually a linear function. To see this, let $v, v' \in V$ and let $\alpha \in \mathbf{F}$. Then

$$T_w(v + v') = \langle T(v + v'), w \rangle_W = \langle T(v), w \rangle_W + \langle T(v'), w \rangle_W = T_w(v) + T_w(v').$$
$$T_w(\alpha v) = \langle T(\alpha v), w \rangle_W = \alpha \langle T(v), w \rangle_W = \alpha T_w(v).$$

Definition 5.5.7 (Adjoint). Since $T_w \colon V \to \mathbf{F}$ is a linear functional, we can apply Theorem 5.5.6 to get a unique vector in V, which we denote by $T^*(w)$, such that for all $v \in V$,

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

As we will see shortly, T^* is a linear transformation from W to V, which we call the **adjoint** of T. Also, recalling the definition of T_w, we have

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

Remark 5.5.8. Note that $T \colon V \to W$, whereas $T^* \colon W \to V$.

Remark 5.5.9. We have added subscripts to the above inner products to emphasize that the inner product in W could be different from the inner product in V. From now on, we will drop these subscripts.

Lemma 5.5.10. *Let $T\colon V \to W$ be a linear transformation between finite-dimensional inner product spaces. Then $T^*\colon W \to V$ is a linear transformation*

Proof. Let $w, w' \in W$ and let $\alpha \in \mathbf{F}$. We will first show that $T^*(w+w') = T^*(w) + T^*(w')$. By the definition of T^*, for all $v \in V$,

$$\langle T(v), w+w'\rangle = \langle v, T^*(w+w')\rangle.$$

So, rearranging things and applying the definition of T^* again,

$$\langle v, T^*(w+w')\rangle = \langle T(v), w\rangle + \langle T(v), w'\rangle = \langle v, T^*(w)\rangle + \langle v, T^*(w')\rangle = \langle v, T^*(w) + T^*(w')\rangle.$$

By the uniqueness part of the Riesz Representation Theorem (Theorem 5.5.6), we therefore have $T^*(w+w') = T^*(w) + T^*(w')$.

We now show that $T^*(\alpha w) = \alpha T^*(w)$.

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

By the uniqueness part of the Riesz Representation Theorem (Theorem 5.5.6), we therefore have $T^*(\alpha w) = \alpha T^*(w)$. □

Definition 5.5.11 (**Adjoint of a Matrix**)**.** Let A be an $m \times n$ matrix with $A_{jk} \in \mathbf{C}$, $1 \le j \le m$, $1 \le k \le n$. The **adjoint** of A, denoted by $A^\dagger$, is an $n \times m$ matrix with entries $(A^\dagger)_{jk} := \overline{A_{kj}}$, $1 \le j \le n$, $1 \le k \le m$.

Theorem 5.5.12. *Let $T\colon V \to W$ be a linear transformation between inner product spaces V, W. Let $\beta = (v_1, \ldots, v_n)$ be an orthonormal basis of V and let $\gamma = (w_1, \ldots, w_m)$ be an orthonormal basis of W. Then*

$$[T^*]_\gamma^\beta = ([T]_\beta^\gamma)^\dagger.$$

Proof. Let $w \in W$. From Corollary 5.3.11, recall that

$$[w]^\gamma = \begin{pmatrix} \langle w, w_1\rangle \\ \vdots \\ \langle w, w_m\rangle \end{pmatrix}.$$

Also, from Remark 5.3.12, $[T]_\beta^\gamma$ has columns $[T(v_1)]^\gamma, \ldots, [T(v_n)]^\gamma$. That is,

$$[T]_\beta^\gamma = \begin{pmatrix} \langle T(v_1), w_1\rangle & \langle T(v_2), w_1\rangle & \cdots & \langle T(v_n), w_1\rangle \\ \langle T(v_1), w_2\rangle & \langle T(v_2), w_2\rangle & \cdots & \langle T(v_n), w_2\rangle \\ \vdots & \vdots & \cdots & \vdots \\ \langle T(v_1), w_m\rangle & \langle T(v_2), w_m\rangle & \cdots & \langle T(v_n), w_m\rangle \end{pmatrix}.$$

Similarly, $[T^*]_\gamma^\beta$ is an $n \times m$ matrix with (i,j) entry $\langle T^*(w_j), v_i\rangle$. And

$$\langle T^*(w_j), v_i\rangle = \overline{\langle v_i, T^*(w_j)\rangle} = \overline{\langle T(v_i), w_j\rangle}.$$

That is,

$$[T^*]_\gamma^\beta = \begin{pmatrix} \overline{\langle T(v_1), w_1\rangle} & \overline{\langle T(v_1), w_2\rangle} & \cdots & \overline{\langle T(v_1), w_m\rangle} \\ \overline{\langle T(v_2), w_1\rangle} & \overline{\langle T(v_2), w_2\rangle} & \cdots & \overline{\langle T(v_2), w_m\rangle} \\ \vdots & \vdots & \cdots & \vdots \\ \overline{\langle T(v_n), w_1\rangle} & \overline{\langle T(v_n), w_2\rangle} & \cdots & \overline{\langle T(v_n), w_m\rangle} \end{pmatrix}.$$

In conclusion $[T^*]_\gamma^\beta = ([T]_\beta^\gamma)^\dagger$. □

Corollary 5.5.13. *Let* $\mathbf{F}$ *denote* $\mathbf{R}$ *or* $\mathbf{C}$*. Let* A *be an* $m \times n$ *matrix with elements in* $\mathbf{F}$*. Let* $\mathbf{F}^n$ *and* $\mathbf{F}^m$ *respectively denote the usual vector spaces* $\mathbf{F}^n$ *and* $\mathbf{F}^m$ *with their standard inner products. Then the adjoint of* $L_A \colon \mathbf{F}^n \to \mathbf{F}^m$ *is* $L_{A^\dagger}$*.*

Proof. Note that $L_A \colon \mathbf{F}^m \to \mathbf{F}^n$ and $L_{A^\dagger} \colon \mathbf{F}^n \to \mathbf{F}^m$. Let β be the standard basis of $\mathbf{F}^m$ and let γ be the standard basis of $\mathbf{F}^n$. From Theorem 5.5.12,

$$[L_A^*]_\gamma^\beta = ([L_A]_\beta^\gamma)^\dagger = A^\dagger = [L_{A^\dagger}]_\gamma^\beta.$$

So, $L_A^* = L_{A^\dagger}$, as desired. □

Remark 5.5.14. Let $\mathbf{F}$ denote $\mathbf{R}$ or $\mathbf{C}$. Let A be an $m \times n$ matrix with elements in $\mathbf{F}$. Then for any $v \in \mathbf{F}^n$ and for any $w \in \mathbf{F}^m$,

$$\langle Av, w\rangle = \langle v, A^\dagger w\rangle.$$

In this equality, the inner product on the left is the standard inner product on $\mathbf{F}^m$, and the inner product on the right is the standard inner product on $\mathbf{F}^n$. Note that if we change the inner product, then the adjoint could possibly change as well. For example, suppose $n = 2$ and we use the inner product

$$\langle (v_1, v_2), (w_1, w_2)\rangle' := (v_1, v_2) \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix} (w_1, w_2)^t, \qquad (v_1, v_2), (w_1, w_2) \in \mathbf{R}^2.$$

Then it is not true that $\langle Av, w\rangle' = \langle v, A^\dagger w\rangle'$. For example, choose $v = (1, 0)$, $w = (1, 0)$, $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Then $\langle Av, w\rangle' = \langle (1, 0), (1, 0)\rangle' = (1, 0)(1, 1/2)^t = 1$, while $\langle v, A^\dagger w\rangle' = \langle (1, 0), (1, 1)\rangle' = (1, 0)(3/2, 3/2)^t = 3/2$.

Exercise 5.5.15. Find the adjoint of $L_A \colon \mathbf{R}^2 \to \mathbf{R}^2$ in the above example, where $\mathbf{R}^2$ is equipped with the inner product $\langle\, , \rangle'$.

Exercise 5.5.16. Let $\mathbf{F}$ denote $\mathbf{R}$ or $\mathbf{C}$. Let $T \colon V \to W$, $S \colon V \to W$ and let $R \colon U \to V$ be linear transformations between inner product spaces U, V, W over $\mathbf{F}$. Verify the following facts

(a) $(T + S)^* = T^* + S^*$.
(b) For all $\alpha \in \mathbf{F}$, $(\alpha T)^* = \overline{\alpha} T^*$.
(c) $(T^*)^* = T$.
(d) $(TR)^* = R^* T^*$.
(e) If T is invertible, then $(T^{-1})^* = (T^*)^{-1}$.

Exercise 5.5.17. Let A be an $m \times n$ matrix. Show that $\operatorname{rank}(A) = \operatorname{rank}(A^\dagger)$.

Exercise 5.5.18. Let A be an $n \times n$ matrix with elements in $\mathbf{C}$. Then $\det(A^\dagger) = \overline{\det(A)}$.

5.6. **Normal Operators.** One of the ultimate goals of this course is to take an arbitrary linear transformation and either diagonalize it, or show that it cannot be diagonalized. Such a result provides the starting point for many further investigations. We cannot fully realize this goal in this course. However, we will identity a large class of linear transformations (i.e. operators) that can be diagonalized, and that appear often in practice. Two such classes of operators are normal and self-adjoint operators. This course will conclude by showing that these two classes of operators can be diagonalized. Such a diagonalization result is referred to as a **spectral theorem**. The **spectrum** of a linear operator is its set of

eigenvalues. This terminology may seem a bit strange, since we usually refer to the spectrum of an electromagnetic wave. However, this conflation of terminology is no coincidence. For example, the spectral theorem for infinite-dimensional vector spaces (which is outside the scope of this course) demonstrates mathematically the discreteness of the energy emissions of the hydrogen atom. In particular, there is a self-adjoint operator whose set of eigenvalues is the energy emission spectrum of the hydrogen atom. And the eigenvectors give the (infinite set of) atomic orbitals that you learned in chemistry class, the first of which you called s,p,d and f orbitals. (Beware: in infinite-dimensional spaces, self-adjointness becomes more complicated than in the finite-dimensional case.)

Definition 5.6.1 (**Normal Operator**)**.** Let V be a finite-dimensional inner product space. Let $T\colon V \to V$ be a linear transformation. Recall that $T^*\colon V \to V$ is also a linear transformation. We say that T is a **normal operator** if $TT^* = T^*T$.

Example 5.6.2. Define $T\colon \mathbf{R}^2 \to \mathbf{R}^2$ by $T(x, y) = (y, -x)$ for all $x, y \in \mathbf{R}$. Then $T^*(x, y) = (-y, x)$, by the definition of T^*. Observe,

$$TT^*(x, y) = T(-y, x) = (x, y).$$
$$T^*T(x, y) = T^*(y, -x) = (x, y).$$

So, $T^*T = TT^*$, so T is normal.

Definition 5.6.3 (**Normal Matrix**)**.** Let A be an $n \times n$ matrix. We say that A is **normal** if $AA^\dagger = A^\dagger A$.

Example 5.6.4. Every diagonal matrix is normal.

Proposition 5.6.5. *Let $T\colon V \to V$ be a linear transformation on a finite-dimensional inner product space V. Let β be an orthonormal basis of V. Then $T\colon V \to V$ is normal if and only if $[T]_\beta^\beta$ is normal.*

Proof. Suppose T is normal. Then $TT^* = T^*T$. Taking the matrix representation of this identity,

$$[T]_\beta^\beta [T^*]_\beta^\beta = [TT^*]_\beta^\beta = [T^*T]_\beta^\beta = [T^*]_\beta^\beta [T]_\beta^\beta.$$

From Theorem 5.5.12, $[T^*]_\beta^\beta$ is the adjoint of $[T]_\beta^\beta$. So, $[T]_\beta^\beta$ is normal.

So, we proved the forward implication. To prove the reverse implication, note that the above steps can be reversed. □

Lemma 5.6.6. *Let $\mathbf{F}$ denote $\mathbf{R}$ or $\mathbf{C}$. Let V be a finite-dimensional inner product space over $\mathbf{F}$. Let $T\colon V \to V$ be a normal operator. Assume that there exists $v \in V$ and $\lambda \in \mathbf{F}$ such that $Tv = \lambda v$. Then $T^*v = \overline{\lambda} v$.*

Proof. It suffices to show that $\|T^*(v) - \overline{\lambda} v\|^2 = 0$. That is, we verify that

$$\langle T^*(v) - \overline{\lambda} v, T^*(v) - \overline{\lambda} v \rangle = 0.$$

Expanding both sides, we equivalently want

$$\langle T^*(v), T^*(v) \rangle - \overline{\lambda} \langle v, T^*(v) \rangle - \lambda \langle T^*(v), v \rangle + |\lambda|^2 \langle v, v \rangle = 0.$$

Using the definition of adjoint and that T is normal, the left-most term is $\langle TT^*v, v \rangle = \langle T^*Tv, v \rangle$. Using the adjoint definition some more, we equivalently want

$$\langle T^*T(v), v \rangle - \overline{\lambda} \langle T(v), v \rangle - \lambda \langle v, T(v) \rangle + |\lambda|^2 \langle v, v \rangle = 0.$$

Using $T(v) = \lambda v$, we equivalently want

$$\lambda\langle T^*(v), v\rangle - |\lambda|^2 \langle v, v\rangle - |\lambda|^2 \langle v, v\rangle + |\lambda|^2 \langle v, v\rangle = 0.$$

Applying the adjoint definition again and simplifying, we want

$$\lambda\langle v, T(v)\rangle - |\lambda|^2 \langle v, v\rangle = 0.$$

Finally, using $T(v) = \lambda v$, we have $\lambda\langle v, T(v)\rangle = |\lambda|^2 \langle v, v\rangle$, completing the proof. □

The following Lemma shows that Proposition 4.5.2 becomes strengthened when T is normal.

Lemma 5.6.7. *Let $T\colon V \to V$ be a normal operator on a finite-dimensional inner product space V. Let v_1, v_2 be two eigenvectors of T with distinct eigenvalues λ_1, λ_2, respectively. Then v_1 is orthogonal to v_2.*

Proof. Since $T(v_1) = \lambda_1 v_1$ and $T(v_2) = \lambda_2 v_2$, we have $T^*(v_1) = \overline{\lambda_1} v_1$ and $T^*(v_2) = \overline{\lambda_2} v_2$ by Lemma 5.6.6. So,

$$\lambda_1\langle v_1, v_2\rangle = \langle T(v_1), v_2\rangle = \langle v_1, T^*(v_2)\rangle = \lambda_2\langle v_1, v_2\rangle.$$

Since $\lambda_1 \neq \lambda_2$, we must have $\langle v_1, v_2\rangle = 0$, as desired. □

Remark 5.6.8. Most linear transformations will not be normal, since they will typically have non-orthogonal eigenvectors.

We now prove a converse to Lemma 5.6.7, which is also a variation on Lemma 4.3.9 for normal T.

Lemma 5.6.9. *Let $T\colon V \to V$ be a linear transformation on a finite-dimensional inner product space V. Suppose β is an orthonormal basis of V consisting of eigenvectors of T. Then T is normal.*

Proof. From Lemma 4.3.9, $[T]_\beta^\beta$ is diagonal. In particular, $[T]_\beta^\beta$ is normal. So, from Proposition 5.6.5, T is normal. □

Theorem 5.6.10 (**The Spectral Theorem for Normal Operators**). *Let $T\colon V \to V$ be a normal operator on a finite-dimensional inner product space V over $\mathbf{C}$. Then there exists an orthonormal basis β of V consisting of eigenvectors of T. In particular, T is diagonalizable.*

Proof. We will prove the theorem by induction on the dimension n of V. Consider first the case $n = 1$. Let β consist of exactly one nonzero unit vector $v \in V$. Since V is one dimensional, for any $w \in V$, there exists $\alpha \in \mathbf{C}$ such that $w = \alpha v$. So, if $T(v) = w$ for some $w \in V$, we have $T(v) = \alpha v$, so that v is an eigenvector of T. In conclusion, the theorem holds for $n = 1$.

Now, suppose the theorem holds for a fixed $n \geq 1$, and consider the case $\dim(V) = n + 1$. Let $f(\lambda)$ be the characteristic polynomial of some matrix representation of T. (Recall that any two matrix representations of T are similar by Lemma 2.7.3, and two similar matrices have the same characteristic polynomial by Lemma 4.4.3. So, the matrix representation that we use for T does not affect f.) From the Fundamental Theorem of Algebra (Theorem 4.4.12), f has $n + 1$ zeros. In particular, f has one zero. So, T has at least one eigenvalue $\lambda_1 \in \mathbf{C}$, and at least one eigenvector $v_1 \in V$, $v_1 \neq 0$ with $T(v_1) = \lambda_1 v_1$. Replacing v_1 with $v_1 / \|v_1\|$ if necessary, we may assume that $\|v_1\| = 1$.

Since $T(v_1) = \lambda_1 v_1$, Lemma 5.6.6 shows that $T^*(v_1) = \overline{\lambda_1} v_1$. Let $W := \{av_1 \colon a \in \mathbf{C}\}$ denote the span of v_1. Observe that $W \subseteq V$ is a one-dimensional subspace. Let $W^\perp := \{v \in V \colon \langle v, v_1\rangle = 0\}$ denote the orthogonal complement of W. Recall that $W^\perp$ is a subspace of V by Exercise 5.4.15, and $\dim(W^\perp) = n + 1 - 1 = n$ by Corollary 5.4.18.

We would like to apply the inductive hypothesis to T, where we restrict the domain of T to the subspace $W^\perp$. In order for the inductive hypothesis to apply, we need to show that the restriction of T to $W^\perp$ satisfies the hypotheses of the theorem. That is, we need to show:

(a) $T(W^\perp) \subseteq W^\perp$, i.e. that $W^\perp$ is invariant under T.
(b) $T^*(W^\perp) \subseteq W^\perp$.
(c) T and T^* are adjoints of each other when we consider them as operators on $W^\perp$.

Proof of (a). Let $w \in W^\perp$, so that $\langle w, v_1\rangle = 0$. Then

$$0 = \lambda_1 \langle w, v_1\rangle = \langle w, T^*(v_1)\rangle = \langle T(w), v_1\rangle.$$

So, $T(w) \in W^\perp$, as desired.

Proof of (b). Let $w \in W^\perp$, so that $\langle w, v_1\rangle = 0$. Then

$$0 = \overline{\lambda_1} \langle w, v_1\rangle = \langle w, T(v_1)\rangle = \langle T^*(w), v_1\rangle.$$

So, $T^*(w) \in W^\perp$, as desired.

Proof of (c). Let $v, w \in W^\perp$. We need to show that there exists $x \in W^\perp$ such that

$$\langle T(v), w\rangle = \langle v, x\rangle. \qquad (*)$$

Since T^* is the adjoint of T, we know that $x := T^*(w)$ is the unique vector in V such that $(*)$ holds, by the Riesz Representation Theorem (Theorem 5.5.6). So, we need to show that $T^*(w) \in W^\perp$. But this follows from part (b).

Having proven parts (a),(b) and (c), we can finally apply the inductive hypothesis to T, where we restrict the domain of T to $W^\perp$. That is, there exists an orthonormal basis $(v_2, \ldots, v_{n+1})$ of $W^\perp$ consisting of eigenvectors of T. Since $v_1 \in W$, v_1 is orthogonal to the vectors $v_2, \ldots, v_{n+1}$. So, the set of vectors $v_1, \ldots, v_{n+1}$ is an orthonormal set (recalling $\|v_1\| = 1$). Since $\dim(V) = n + 1$, Corollary 5.3.9 says $v_1, \ldots, v_{n+1}$ is a basis of V, as desired. □

5.7. **Self-Adjoint Operators.**

Definition 5.7.1 (**Self-Adjoint Operator**)**.** Let $\mathbf{F}$ denote $\mathbf{R}$ or $\mathbf{C}$. Let V be a finite-dimensional inner product space over $\mathbf{F}$. Let $T \colon V \to V$ be a linear transformation. Then T is called a **self-adjoint operator** if $T^* = T$. A square matrix A is said to be **self-adjoint** if $A = A^\dagger$.

Remark 5.7.2. Let $T \colon V \to V$ be a linear transformation on a finite-dimensional inner product space V. If T is self-adjoint, then T is normal. But if T is normal, then T is not necessarily self-adjoint.

Example 5.7.3. The linear transformation $T \colon \mathbf{R}^2 \to \mathbf{R}^2$ defined by $T(x, y) = (y, -x)$ is normal but not self-adjoint, since it has adjoint $T^*(x, y) = (-y, x)$. However, the linear transformation $T \colon \mathbf{R}^2 \to \mathbf{R}^2$ defined by $T(x, y) = (y, x)$ is self-adjoint.

Remark 5.7.4. Let $T \colon V \to V$ be a linear transformation on a finite-dimensional inner product space V over $\mathbf{F}$. If T is self-adjoint and if $\mathbf{F} = \mathbf{C}$, then T is sometimes called **Hermitian**. If T is self-adjoint and if $\mathbf{F} = \mathbf{R}$, then T is **symmetric**. A square complex

matrix A with $A = A^\dagger$ is also called **Hermitian**. And a square real matrix with $A = A^\dagger$ is called **symmetric**, since $A = A^\dagger$ becomes $A = A^t$.

Theorem 5.7.5. *Let* $\mathbf{F}$ *denote* $\mathbf{R}$ *or* $\mathbf{C}$*. Let* V *be a finite-dimensional inner product space over* $\mathbf{F}$*. Let* $T\colon V \to V$ *be a self-adjoint linear transformation. Then all eigenvalues of* T *are real.*

Proof. Let $\lambda \in \mathbf{C}$ be any eigenvalue of T. So, there exists $v \in V$ with $v \neq 0$ such that $T(v) = \lambda v$. Lemma 5.6.6 shows that $T^*(v) = \overline{\lambda} v$. Since $T = T^*$, we conclude that $\lambda = \overline{\lambda}$, so that $\lambda \in \mathbf{R}$, as desired. □

Remark 5.7.6. Similarly, all eigenvalues of a Hermitian matrix are real.

Proposition 5.7.7. *Let* $T\colon V \to V$ *be a linear transformation on a finite-dimensional inner product space* V*. Let* β *be an orthonormal basis of* V*. Then* $T\colon V \to V$ *is self-adjoint if and only if* $[T]_\beta^\beta$ *is self-adjoint.*

Proof. Suppose T is self-adjoint. Then $T = T^*$. From Theorem 5.5.12, $[T^*]_\beta^\beta$ is the adjoint of $[T]_\beta^\beta$. So, $[T]_\beta^\beta$ is self-adjoint. We proved the forward implication. To prove the reverse implication, note that the above steps can be reversed. □

Corollary 5.7.8. *Let* A *be an* $n \times n$ *complex Hermitian matrix, and let* $f(\lambda) := \det(A - \lambda I)$ *be the characteristic polynomial of* A*. Then there exist* $\lambda_1, \ldots, \lambda_n \in \mathbf{R}$ *such that* $f(\lambda) = \prod_{i=1}^n (\lambda_i - \lambda)$.

Proof. From the Fundamental Theorem of Algebra (Theorem 4.4.12), there exist $\lambda_0, \ldots, \lambda_n \in \mathbf{C}$ such that $f(\lambda) = \lambda_0 \prod_{i=1}^n (\lambda_i - \lambda)$. Recall that coefficient of the degree n term of $f(\lambda)$ is $(-1)^n$ by Theorem 4.4.8. So, $\lambda_0 = 1$. From Remark 5.7.6, $\lambda_i \in \mathbf{R}$ for all $i \in \{1, \ldots, n\}$. □

Lemma 5.7.9. *Let* $\mathbf{F}$ *denote either* $\mathbf{R}$ *or* $\mathbf{C}$*. Let* $T\colon V \to V$ *be a linear transformation on a finite-dimensional inner product space* V *over* $\mathbf{F}$*. Suppose there exists an orthonormal basis* β *of* V *consisting of eigenvectors of* T *with real eigenvalues. Then* T *is self-adjoint.*

Proof. From Lemma 4.3.9, $[T]_\beta^\beta$ is diagonal with real entries. In particular, $[T]_\beta^\beta$ is self-adjoint So, from Proposition 5.7.7, T is self-adjoint. □

Theorem 5.7.10 (The Spectral Theorem for Self-Adjoint Operators). *Let* $\mathbf{F}$ *denote either* $\mathbf{R}$ *or* $\mathbf{C}$*. Let* $T\colon V \to V$ *be a self-adjoint operator on a finite-dimensional inner product space* V *over* $\mathbf{F}$*. Then there exists an orthonormal basis* β *of* V *consisting of eigenvectors of* T*. In particular,* T *is diagonalizable. Moreover, all eigenvalues of* T *are real.*

Proof. Since T is self-adjoint, T is normal. So, if $\mathbf{F} = \mathbf{C}$, the result follows directly from the Spectral Theorem for normal operators (Theorem 5.6.10). Then we apply Theorem 5.7.5 to finish. If $\mathbf{F} = \mathbf{R}$, we repeat the proof of Theorem 5.6.10, replacing $\mathbf{C}$ everywhere by $\mathbf{R}$. The crucial new ingredient in the proof is that, in the inductive step and in the base case of the induction, T has some real eigenvalue $\lambda_1 \in \mathbf{R}$ by Theorem 5.7.5. □

Remark 5.7.11. Note that Theorem 5.6.10 requires V to be a vector space over $\mathbf{C}$. But Theorem 5.7.10 requires V to be a vector space over $\mathbf{R}$ or $\mathbf{C}$. So, every symmetric operator on a real inner product space is diagonalizable.

Remark 5.7.12. In conclusion, self-adjoint operators are really nice, since they have an orthonormal basis of eigenvectors (so they can be diagonalized), and all of their eigenvalues are real.

5.8. **Orthogonal and Unitary Operators (Bonus Section).**

Definition 5.8.1 (**Unitary Operators**)**.** Let $\mathbf{F}$ denote $\mathbf{R}$ or $\mathbf{C}$. Let V be a finite-dimensional inner product space over $\mathbf{F}$. Let $T\colon V \to V$ be a linear transformation. Then T is called a **unitary operator** if $TT^* = T^*T = I_V$. A square matrix A is called **unitary** if $AA^\dagger = A^\dagger A = I$.

Remark 5.8.2. Let $T\colon V \to V$ be a linear transformation on a finite-dimensional inner product space V. If T is unitary, then T is normal.

Remark 5.8.3. Let $T\colon V \to V$ be a linear transformation on a finite-dimensional inner product space V over $\mathbf{F}$. If T is unitary and $\mathbf{F} = \mathbf{R}$, then T is called **orthogonal**. A square real matrix A with $AA^\dagger = A^\dagger A = I$ is also called **orthogonal**, since we then have $AA^t = A^t A = I$.

Theorem 5.8.4. *Let $\mathbf{F}$ denote $\mathbf{R}$ or $\mathbf{C}$. Let V be a finite-dimensional inner product space over $\mathbf{F}$. Let $T\colon V \to V$ be a unitary operator. Then all eigenvalues of T have absolute value 1.*

Proof. Let $\lambda \in \mathbf{C}$ be any eigenvalue of T. So, there exists $v \in V$ with $v \neq 0$ such that $T(v) = \lambda v$. Lemma 5.6.6 shows that $T^*(v) = \overline{\lambda} v$. So, using $T^*T = I_V$,

$$|\lambda|^2 \langle v, v\rangle = \langle \lambda v, \lambda v\rangle = \langle T(v), T(v)\rangle = \langle T^*T(v), v\rangle = \langle v, v\rangle.$$

Since $v \neq 0$, we conclude that $|\lambda|^2 = 1$, as desired. □

Remark 5.8.5. Similarly, all eigenvalues of a unitary matrix have absolute value 1.

Proposition 5.8.6. *Let $T\colon V \to V$ be a linear transformation on a finite-dimensional inner product space V. Let β be an orthonormal basis of V. Then $T\colon V \to V$ is unitary if and only if $[T]_\beta^\beta$ is unitary.*

Proof. Suppose T is unitary. Then $TT^* = T^*T = I_V$. Taking the matrix representation,

$$[T]_\beta^\beta [T^*]_\beta^\beta = [TT^*]_\beta^\beta = [T^*T]_\beta^\beta = [T^*]_\beta^\beta [T]_\beta^\beta = [I_V]_\beta^\beta = I.$$

From Theorem 5.5.12, $[T^*]_\beta^\beta$ is the adjoint of $[T]_\beta^\beta$. So, $[T]_\beta^\beta$ is unitary. We proved the forward implication. To prove the reverse implication, note that the above steps can be reversed. □

Corollary 5.8.7. *Let A be an $n \times n$ unitary matrix, and let $f(\lambda) := \det(A - \lambda I)$ be the characteristic polynomial of A. Then there exist $\lambda_1, \ldots, \lambda_n \in \mathbf{C}$ with $|\lambda_i| = 1$ for all $i \in \{1, \ldots, n\}$ such that $f(\lambda) = \prod_{i=1}^n (\lambda_i - \lambda)$.*

Proof. From the Fundamental Theorem of Algebra (Theorem 4.4.12), there exist $\lambda_0, \ldots, \lambda_n \in \mathbf{C}$ such that $f(\lambda) = \lambda_0 \prod_{i=1}^n (\lambda_i - \lambda)$. Recall that coefficient of the degree n term of $f(\lambda)$ is $(-1)^n$ by Theorem 4.4.8. So, $\lambda_0 = 1$. From Remark 5.8.5, $|\lambda_i| = 1$ for all $i \in \{1, \ldots, n\}$. □

Lemma 5.8.8. *Let $\mathbf{F}$ denote either $\mathbf{R}$ or $\mathbf{C}$. Let $T\colon V \to V$ be a linear transformation on a finite-dimensional inner product space V over $\mathbf{F}$. Suppose there exists an orthonormal basis β of V consisting of eigenvectors of T with eigenvalues of absolute value 1. Then T is unitary.*

Proof. From Lemma 4.3.9, $[T]_\beta^\beta$ is diagonal with entries of absolute value 1. In particular, $[T]_\beta^\beta$ is unitary. So, from Proposition 5.8.6, T is unitary. □

Theorem 5.8.9 (**The Spectral Theorem for Unitary Operators**). *Let $T\colon V \to V$ be a unitary operator on a finite-dimensional inner product space V over $\mathbf{C}$. Then there exists an orthonormal basis β of V consisting of eigenvectors of T. In particular, T is diagonalizable. Moreover, all eigenvalues of T have absolute value* 1.

Proof. Since T is unitary, T is normal. So, the result follows directly from the Spectral Theorem for normal operators (Theorem 5.6.10). Then we apply Theorem 5.8.4 to finish. □

Remark 5.8.10. Note that Theorem 5.8.9 requires V to be a vector space over $\mathbf{C}$. In the case that V is a vector space over $\mathbf{R}$, the corresponding spectral theorem becomes very restricted, since the only real numbers with absolute value one are 1 and -1. So, if we want to diagonalize an orthogonal operator over $\mathbf{R}$, T must have all eigenvalues 1 or -1. Even though we can diagonalize an orthogonal operator over $\mathbf{C}$ by Theorem 5.8.9, we can essentially never diagonalize an orthogonal operator over $\mathbf{R}$. Nevertheless, let's present the result for diagonalization over $\mathbf{R}$.

If $\mathbf{F} = \mathbf{R}$, and if T is both self-adjoint and unitary, then the Spectral Theorem for self-adjoint operators (Theorem 5.7.10) together with Theorem 5.8.4 show: there exists an orthonormal basis β of V consisting of eigenvectors of T. So, T is diagonalizable, and all eigenvalues of T are 1 or -1. Conversely, suppose T is a linear operator on an inner product space V over $\mathbf{R}$, and suppose there exists a basis β of V consisting of eigenvectors of T with eigenvalues 1 or -1. Then $[T]_\beta^\beta$ is orthogonal. So T is orthogonal by Proposition 5.8.6. By Lemma 5.7.9, T is also self-adjoint.

6. Appendix: Notation

Let A, B be sets in a space X. Let m, n be a nonnegative integers. Let $\mathbf{F}$ be a field.

$\mathbf{Z} := \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$, the integers

$\mathbf{N} := \{0, 1, 2, 3, 4, 5, \ldots\}$, the natural numbers

$\mathbf{Q} := \{m/n \colon m, n \in \mathbf{Z}, n \neq 0\}$, the rationals

$\mathbf{R}$ denotes the set of real numbers

$\mathbf{C} := \{x + y\sqrt{-1} : x, y \in \mathbf{R}\}$, the complex numbers

$\overline{x + y\sqrt{-1}} := x - y\sqrt{-1}, \quad x, y \in \mathbf{R}$, the complex conjugate

$\emptyset$ denotes the empty set, the set consisting of zero elements

$\in$ means "is an element of." For example, $2 \in \mathbf{Z}$ is read as "2 is an element of $\mathbf{Z}$."

$\forall$ means "for all"

$\exists$ means "there exists"

$\mathbf{F}^n := \{(x_1, \ldots, x_n) \colon x_i \in \mathbf{F}, \forall\, i \in \{1, \ldots, n\}\}$

$A \subseteq B$ means $\forall\ a \in A$, we have $a \in B$, so A is contained in B

$A \smallsetminus B := \{x \in A \colon x \notin B\}$

$A^c := X \smallsetminus A$, the complement of A

$A \cap B$ denotes the intersection of A and B

$A \cup B$ denotes the union of A and B

$C(\mathbf{R})$ denotes the set of all continuous functions from $\mathbf{R}$ to $\mathbf{R}$

$P_n(\mathbf{R})$ denotes the set of all real polynomials in one real variable of degree at most n

$P(\mathbf{R})$ denotes the set of all real polynomials in one real variable

$M_{m\times n}(\mathbf{F})$ denotes the vector space of $m \times n$ matrices over the field $\mathbf{F}$

I_n denotes the $n \times n$ identity matrix

det denotes the determinant function

S_n denotes the set of permutations on $\{1, \ldots, n\}$

$\operatorname{sign}(\sigma) := (-1)^N$ where $\sigma \in S_n$ can be written as the composition of N transpositions

Tr denotes the trace function

6.0.1. *Set Theory.* Let V, W be sets, and let $f \colon V \to W$ be a function. Let $X \subseteq V$, $Y \subseteq W$.

$$f(X) := \{f(v) \colon v \in V\}.$$

$$f^{-1}(Y) := \{v \in V \colon f(v) \in Y\}.$$

The function $f \colon V \to W$ is said to be **injective** (or **one-to-one**) if: for every $v, v' \in V$, if $f(v) = f(v')$, then $v = v'$.

The function $f \colon V \to W$ is said to be **surjective** (or **onto**) if: for every $w \in W$, there exists $v \in V$ such that $f(v) = w$.

The function $f\colon V \to W$ is said to be **bijective** (or a **one-to-one correspondence**) if: for every $w \in W$, there exists exactly one $v \in V$ such that $f(v) = w$. A function $f\colon V \to W$ is bijective if and only if it is both injective and surjective.

Two sets X, Y are said to have the same **cardinality** if there exists a bijection from V onto W.

The **identity map** $I\colon X \to X$ is defined by $I(x) = x$ for all $x \in X$. To emphasize that the domain and range are both X, we sometimes write I_X for the identity map on X. Let $f\colon X \to X$. We write f^2 to denote f composed with itself: $f \circ f$. More generally, for any $n \in \mathbf{N}$, we write f^n to denote f composed with itself n times: $f \circ f \circ \cdots \circ f$.

Let V, W be vector spaces over a field $\mathbf{F}$. Then $\mathcal{L}(V, W)$ denotes the set of linear transformations from V to W, and $\mathcal{L}(V)$ denotes the set of linear transformations from V to V. Let $T\colon V \to W$ be a linear transformation between inner product spaces. Then $T^*\colon W \to V$ denotes the adjoint of T.

UCLA Department of Mathematics, Los Angeles, CA 90095-1555
E-mail address: heilman@math.ucla.edu