Notes for Linear Algebra

Jonathan Cui

Ver. 20231204

1 Preliminaries

Definition 1.1. A field is a set $\mathbb{F}$ equipped with two operations $+, \cdot : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ that satisfy the following: There exist $0 \in \mathbb{F}$ and $1 \in \mathbb{F}$ such that

- (Commutativity) $\forall a, b \in \mathbb{F}, a + b = b + a \wedge a \cdot b = b \cdot a$;
- (Associativity) $\forall a, b, c \in \mathbb{F}, (a + b) + c = a + (b + c) \wedge (a \cdot b) \cdot c = a \cdot (b \cdot c)$;
- (Identity) $\forall a \in \mathbb{F}, a + 0 = a \cdot 1 = a$;
- (Inverse) $\forall a \in \mathbb{F}, (\exists b \in \mathbb{F}, a + b = 0) \wedge (a \neq 0 \implies \exists b \in \mathbb{F}, a \cdot b = 1)$;
- (Distributivity) $\forall a, b, c \in \mathbb{F}, a \cdot (b + c) = a \cdot b + a \cdot c$.

By the way, some people explicitly state the closure property of $+$ and $\cdot$, but we don't need that since we already said $+, \cdot : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$.

Before we introduce vectors, we'll define some structures to work with.

Definition 1.2. Given a field $\mathbb{F}$, the set $\mathbb{F}^n$ is defined as the collection of all n-tuples with components in $\mathbb{F}$; that is,

$$\mathbb{F}^n := \{(x_1, \cdots, x_n) \mid x_1, \cdots, x_n \in \mathbb{F}\}.$$

We equip the set with two operations, $+, \cdot : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}^n$, defined as follows:

$$\begin{aligned}(x_1, \cdots, x_n) + (y_1, \cdots, y_n) &:= (x_1 + y_1, \cdots, x_n + y_n),\\ c \cdot (x_1, \cdots, x_n) &:= (c \cdot x_1, \cdots, c \cdot x_n),\end{aligned}$$

where $c, x_1, \cdots, x_n, y_1, \cdots, y_n \in \mathbb{F}$.

Definition 1.3. Given a field $\mathbb{F}$, the set $\mathcal{P}(\mathbb{F}) \subset \mathbb{F}^{\mathbb{F}}$ is defined as the collection of all polynomials from $\mathbb{F}$ to $\mathbb{F}$; that is,

$$\mathcal{P}(\mathbb{F}) := \bigcup_{n=0}^{\infty} \{\lambda x.(a_0 + a_1 x + \cdots + a_n x^n) \mid a_0, \cdots, a_n \in \mathbb{F}\}.$$

We equip the set with two operations, $+, \cdot : \mathcal{P}(\mathbb{F}) \times \mathcal{P}(\mathbb{F}) \to \mathcal{P}(\mathbb{F})$, defined as follows:

$$\begin{aligned}f + g &:= \lambda x.(f(x) + g(x)),\\ c \cdot f &:= \lambda x.(c \cdot f(x)),\end{aligned}$$

where $f, g \in \mathcal{P}(\mathbb{F})$ and $c \in \mathbb{F}$.

A nonzero polynomial $f \in \mathcal{P}(\mathbb{F}) \backslash \{\lambda x.0\}$ is said to have degree n if there exists $a_0, \cdots, a_n \in \mathbb{F}$ with $a_n \neq 0$ such that

$$\forall x \in \mathbb{F}, f(x) = a_0 + a_1 x + \cdots + a_n x^n.$$

The zero polynomial $\lambda x.0$ is said to have degree negative infinity. Note that all polynomials have a unique degree, which then defines the function $\deg\colon \mathcal{P}(\mathbb{F}) \to \{-\infty\} \cup \mathbb{Z}_{\geq 0}$ that maps a polynomial to its degree.

Given an arbitrary $n \in \mathbb{Z}_{\geq 0}$, the set $\underline{\mathcal{P}_n(\mathbb{F})}$ is defined as the subset of $\mathcal{P}(\mathbb{F})$ of all polynomials with degree at most n. Clearly, $\mathcal{P}(\mathbb{F})$ and any $\mathcal{P}_n(\mathbb{F})$ are vector spaces.

2 Vector Spaces

Definition 2.1. A vector space over a field $\mathbb{F}$ is a set V equipped with two operations $+\colon V \times V \to V$ and $\cdot\colon \mathbb{F} \times V \to V$ that satisfy the following: there exist $0 \in V$ such that

- $\forall u, v \in V, u + v = v + u$;
- $\forall u, v, w \in V, (u + v) + w = u + (v + w)$;
- $\forall v \in V, v + 0 = v$;
- $\forall v \in V, \exists w \in V, v + w = 0$;
- $\forall v \in V, 1 \cdot v = v$;
- $\forall a, b \in \mathbb{F}, \forall v \in V, a \cdot (b \cdot v) = (a \cdot b) \cdot v$;
- $\forall a, b \in \mathbb{F}, \forall v \in V, (a + b) \cdot v = a \cdot v + b \cdot v$;
- $\forall a \in \mathbb{F}, \forall u, v \in V, a \cdot (u + v) = a \cdot u + a \cdot v$.

Note that the symbol 0 can be the additive identity either from the field or from the vector space, which can usually be inferred from context.

Here are some claims.

Proposition 2.2. Given a vector space V over $\mathbb{F}$, the additive identity 0 is unique.

Proof. Suppose $0_1, 0_2 \in V$ are both satisfy the properties of 0 in Definition 2.1. Then,

$$0_1 = 0_1 + 0_2 = 0_2.$$

The proof is finished. □

The use of the symbol 0 without additional setup/condition in a vector space is now justified.

Proposition 2.3. Given a vector space V over $\mathbb{F}$ and an arbitrary $v \in V$, the additive inverse is unique.

Proof. Suppose $w_1, w_2 \in V$ are such that $v + w_1 = v + w_2 = 0$. Then,

$$w_1 + v + w_1 = w_1 + v + w_2 \Rightarrow w_1 = w_2.$$

The proof is complete. □

We therefore introduce the notation $-v$ to represent the additive inverse of $v \in V$, where V is a vector space.

Proposition 2.4. Given a vector space V over $\mathbb{F}$ and an arbitrary $v \in V$, $0 \cdot v = 0$.

Proof. We have $0 \cdot v + 0 \cdot v = (0 + 0) \cdot v = 0 \cdot v$. Then, $-(0 \cdot v) + 0 \cdot v + 0 \cdot v = -(0 \cdot v) + 0 \cdot v$, and hence $0 \cdot v = 0$. □

Proposition 2.5. Given a vector space V over $\mathbb{F}$ and an arbitrary $a \in \mathbb{F}$, $a \cdot 0 = 0$.

Proof. We have $a \cdot 0 + a \cdot 0 = a \cdot (0+0) = a \cdot 0$. Then, $-(a \cdot 0) + a \cdot 0 + a \cdot 0 = -(a \cdot 0) + a \cdot 0$, and hence $a \cdot 0 = 0$. □

Proposition 2.6. Given a vector space V over $\mathbb{F}$ and an arbitrary $v \in V$, $-v = (-1) \cdot v$.

Proof. Observe that $0 = (-1+1) \cdot v = (-1) \cdot v + 1 \cdot v = (-1) \cdot v + v$. By the uniqueness of the additive inverse, we have $-v = (-1) \cdot v$. □

A subspace is a subset that is also a vector space in its own right. Intuitively, it has to have that "linear" structure. For example, any line or plane through the origin is a subspace of $\mathbb{R}^3$.

Definition 2.7. Given a vector space V over $\mathbb{F}$ and a subset $U \subseteq V$, U is said to be a <u>subspace</u> of V if the images $+(U \times U) \subseteq U$ and $\cdot(\mathbb{F} \times U) \subseteq U$.

We will define the sum of subspaces of a vector space, which is really similar to "spanning." For example, any two distinct lines through the origin in $\mathbb{R}^2$ are subspaces whose sum is $\mathbb{R}^2$.

Definition 2.8. Suppose V is a vector space over $\mathbb{F}$ and $U, W \subseteq V$ are subspaces of V. Then, the sum of U and W, denoted as $U + W$, is defined as the set

$$U + W := \{u + w \mid u \in U, w \in W\}.$$

It is clear that the sum of subspaces is commutative and associative. The second property allows us to write things like $U_1 + U_2 + U_3$ without the use of auxiliary parentheses since the resultant sum is always the same regardless of the order of addition.

Proposition 2.9. Suppose V is a vector space over $\mathbb{F}$ and $U, W \subseteq V$ are subspaces of V. Then, $U + W$ is a subspace of V.

Proof. Let $v_1, v_2 \in U + W$ and $c \in \mathbb{F}$. Fix $u_1, u_2 \in U$ and $w_1, w_2 \in W$ such that $v_1 = u_1 + w_1$ and $v_2 = u_2 + w_2$. Then,

$$v_1 + v_2 = u_1 + w_1 + u_2 + w_2 = \underbrace{u_1 + u_2}_{\in U} + \underbrace{w_1 + w_2}_{\in W} \in U + W.$$

Similarly,

$$c \cdot v_1 = c \cdot (u_1 + w_1) = \underbrace{c \cdot u_1}_{\in U} + \underbrace{c \cdot w_1}_{\in W} \in U + W.$$

The proof is complete. □

When we had the earlier example of two lines, we see that they're both necessary in the sense that removing any would cause the sum to "collapse," much like the linear independence of vectors to be introduced later. It's captured in essentially the same way:

Definition 2.10. Suppose V is a vector space over $\mathbb{F}$ and $U_1, \cdots, U_k \subseteq V$ are subspaces of V. Then, $U_1 + \cdots + U_k$ is said to be a direct sum if for any $v \in U_1 + \cdots + U_k$, there exists a unique $(u_1, \cdots, u_k) \in U_1 \times \cdots \times U_k$ such that $v = u_1 + \cdots + u_k$. If so, we represent the sum with the notation $U_1 \oplus \cdots \oplus U_k$, which equals $U_1 + \cdots + U_k$.

The linearity means we can "shift" vectors and their linear combinations simultaneously, so focusing on all vectors can be reduced to focusing on 0, in terms of whether they/it can be represented uniquely. By the way, to make the word "unique" clearer, we mean the unique existence of a tuple in $U_1 \times \cdots \times U_k$ in the following definition.

Proposition 2.11. Suppose V is a vector space over $\mathbb{F}$ and $U_1, \cdots, U_k \subseteq V$ are subspaces of V. Then, the sum $U_1 + \cdots + U_k$ is direct if and only if there exists a unique $(u_1, \cdots, u_k) \in U_1 \times \cdots \times U_k$ such that $u_1 + \cdots + u_k = 0$.

Proof. **"If" direction.** Suppose for the sake of the contrapositive that the sum $U_1 + \cdots + U_k$ is not direct, and thus fix a vector $v \in V$ with representations

$$u_1 + \cdots + u_k = u_1' + \cdots + u_k' = v$$

for $u_1, u_1' \in U_1, \cdots, u_k, u_k' \in U_k$, where $u_i \neq u_i'$ for some $i \in \{1, \cdots, k\}$. Thus,

$$(u_1 - u_1') + \cdots + (u_k - u_k') = 0,$$

where at least one parenthesized factor is nonzero from assumption, which means 0 is not uniquely represented as the sum of vectors from $U_1, \cdots, U_k$ respectively.

"Only if" direction. Suppose $U_1 + \cdots + U_k$ is direct, but $u_1 + \cdots + u_k = u_1' + \cdots + u_k' = 0$ for $u_1, u_1' \in U_1, \cdots, u_k, u_k' \in U_k$, where $u_i \neq u_i'$ for some $i \in \{1, \cdots, k\}$. We thus have

$$(u_1 - u_1') + \cdots + (u_k - u_k') = 0,$$

where at least one parenthesized factor is nonzero from assumption, which means 0 is not uniquely represented as the sum of vectors from $U_1, \cdots, U_k$ respectively. This contradiction completes the proof. □

This proof is a bit crazy since the "if" and the "only if" directions look almost identical. We used the contrapositive for the "if" and proof by contradiction for the "only if." Of course, another way to view this is we can take the negation of both sides of the "if and only if" that we want to prove.

Now the following result again can be seen through the two-lines example: because they intersect only at the origin, there's no "redundancy."

Proposition 2.12. Suppose V is a vector space over $\mathbb{F}$ and $U, W \subseteq V$ are subspaces of V. Then, the sum $U + W$ is direct if and only if $U \cap W = \{0\}$.

Proof. **"If" direction.** Suppose $U + W$ is not direct. Thus, by Proposition 2.11, fix representations

$$u + w = u' + w' = 0$$

for $u, u' \in U$ and $w, w' \in W$, where either $u \neq u'$ or $w \neq w'$. Thus, $(u - u') + (w - w') = 0$, and thus $(w - w')$ is the additive inverse of $(u - u')$; i.e., $(-1)(u - u')$. Hence, $(w - w') \in U$ as well, and thus $(w - w') \in U \cap W$. By symmetry, $(u - u') \in U \cap W$ as well; yet at least one of $(u - u')$ and $(w - w')$ is non-zero.

"Only if" direction. Suppose $U + W$ is direct but $U + W \neq \{0\}$. Fix $v \in U \cap W$ where $v \neq 0$, whose existence is guaranteed as 0 is in all (sub)spaces and hence $0 \in U \cap W$ necessarily. Then, $-v = (-1) \cdot v \in U \cap W$ as well. But $v + (-v) = 0$ where $v \neq 0$, which contradicts the unique 0 element by Proposition 2.11. The proof is now finished. □

3 Finite-Dimensional Vector Spaces

We now define the linear combination of vectors. The language below uses a list, i.e., a tuple, but sets work equally well since the linear combination is only defined for *finitely many* vectors. And of course, we need to take care of the trivial case too. The span comes right after.

Definition 3.1. Given a vector space V over $\mathbb{F}$ and a non-empty list of vectors $u, v_1, \cdots, v_k$ in V, u is said to be a <u>linear combination</u> of $v_1, \cdots, v_k$ if there exists $a_1, \cdots, a_k \in \mathbb{F}$ such that

$$u = a_1 \cdot v_1 + \cdots + a_k \cdot v_k.$$

We also extend the definition so that there is a unique linear combination of no vectors, i.e., 0.

Definition 3.2. Given a vector space V over $\mathbb{F}$ and vectors $v_1, \cdots, v_k \in V$, the span of $v_1, \cdots, v_k$, denoted as $\operatorname{span}(v_1, \cdots, v_k)$, is defined as the collection of all linear combinations of $v_1, \cdots, v_k$; that is,

$$\operatorname{span}(v_1, \cdots, v_k) \coloneqq \{a_1 \cdot v_1 + \cdots + a_k \cdot v_k \mid a_1, \cdots, a_k \in \mathbb{F}\}.$$

If $\operatorname{span}(v_1, \cdots, v_k) = V$, then the list $v_1, \cdots, v_k$ is said to span V. We also say that $v_1, \cdots, v_k$ is a spanning list of V.

This following proposition is just a really intuitive statement. The thing that gets spanned is always a subspace, and the smallest one containing all of the vectors in the spanning list.

Proposition 3.3. Given a vector space V over $\mathbb{F}$ and vectors $v_1, \cdots, v_k \in V$, $\operatorname{span}(v_1, \cdots, v_k)$ is the smallest subspace that contains $\{v_1, \cdots, v_k\}$.

Proof. We first show that the span is indeed a subspace. Let $u, v \in \operatorname{span}(v_1, \cdots, v_k)$ and $c \in \mathbb{F}$ where

$$\begin{aligned} u &= a_1 \cdot v_1 + \cdots + a_k \cdot v_k, \\ v &= a'_1 \cdot v_1 + \cdots + a'_k \cdot v_k. \end{aligned}$$

Then,

$$u + v = \underbrace{(a_1 + a'_1)}_{\in \mathbb{F}} \cdot \cdot v_1 + \cdots + \underbrace{(a_k + a'_k)}_{\in \mathbb{F}} \cdot v_k \in \operatorname{span}(v_1, \cdots, v_k),$$

and

$$c \cdot u = \underbrace{(c \cdot a_1)}_{\in \mathbb{F}} \cdot \cdot v_1 + \cdots + \underbrace{(c \cdot a_k)}_{\in \mathbb{F}} \cdot v_k \in \operatorname{span}(v_1, \cdots, v_k).$$

Now let $U \subseteq V$ be an arbitrary subspace that contains $\{v_1, \cdots, v_k\}$.

Step 1. Since $v_1 \in U$, we have $U \supseteq \{a_1 \cdot v_1 \mid a_1 \in \mathbb{F}\}$.

Step j ($j = 2, \cdots, k$). Suppose steps 1 through $(j-1)$ have been completed, which imply $U \supseteq \{a_1 \cdot v_1 + \cdots + a_{j-1} \cdot v_{j-1} \mid a_1, \cdots, a_{j-1} \in \mathbb{F}\}$. Since $v_j \in U$, we have $a_j \cdot v_j \in U$ for any $a_j \in \mathbb{F}$. Hence, $U \supseteq \{a_1 \cdot v_1 + \cdots + a_j \cdot v_j \mid a_1, \cdots, a_j \in \mathbb{F}\}$.

After completing steps 1 through k, we have shown that U necessarily contains $\operatorname{span}(v_1, \cdots, v_k)$. Hence $\operatorname{span}(v_1, \cdots, v_k)$ is the smallest such subspace. □

When we deal with finitely many vectors and their linear combinations, things get really easy. Recall that the concept of dimension in matrices is defined by the number of vectors that span the space (if it exists). So, we have the following definition:

Definition 3.4. A vector space V over $\mathbb{F}$ is said to be finite dimensional if there exists a finite subset $\{v_1, \cdots, v_k\}$ of V such that $\operatorname{span}(v_1, \cdots, v_k) = V$.

Fact 3.5. Given a field $\mathbb{F}$, $\mathcal{P}(\mathbb{F})$ is not finite dimensional, but for any $n \in \mathbb{Z}_{\geq 0}$, $\mathcal{P}_n(\mathbb{F})$ is finite dimensional.

Proof. Suppose for the sake of contradiction that $f_1, \cdots, f_k \in \mathcal{P}(\mathbb{F})$ span V.

If $f_1 = \cdots = f_k = 0$, then $V = \operatorname{span}(0, \cdots, 0) = \{0\}$. But observe that $\lambda x.x \in \mathcal{P}(\mathbb{F})$, which is not equal to 0 as verifiable by evaluating at $x = 1$. The premise that $f_1, \cdots, f_k$ are all 0 is therefore impossible

Then, let $N \coloneqq \max\{\deg f_1, \cdots, \deg f_k\}$, which is now necessarily an integer. Observe that $\lambda x.x^{N+1} \in \mathcal{P}(\mathbb{F})$, but it is not in the span of $f_1, \cdots, f_k$, a contradiction.

Observe also that for any $n \in \mathbb{Z}_{\geq 0}$, $\operatorname{span}(\lambda x.1, \cdots, \lambda x.x^n) = \mathcal{P}_n(\mathbb{F})$. □

Now let's tackle linear dependence and independence.

Definition 3.6. Given a vector space V over $\mathbb{F}$, a list of vectors $v_1, \cdots, v_k$ in V is said to be linearly independent if for any $a_1, \cdots, a_k \in \mathbb{F}$ such that

$$a_1 \cdot v_1 + \cdots + a_k \cdot v_k = 0,$$

we have $a_1 = \cdots, a_k = 0$. We define the empty list of vectors to be linearly independent trivially.

If a list of vectors is not linearly independent, it is said to be linearly dependent.

A list of two vectors is linearly dependent if and only if one vector is a scalar multiple of the other. In general, a list of vectors is linearly dependent if and only if one of the vectors is a linear combination of the rest (including the empty list).

Lemma 3.7 (Linear Dependence)**.** Suppose V is a vector space over $\mathbb{F}$ and let $v_1, \cdots, v_k$ be a linearly dependent list of vectors in V. Then, there exists $j \in \{1, \cdots, k\}$ such that

- $v_j \in \operatorname{span}(v_1, \cdots, v_{j-1})$ and
- $\operatorname{span}(v_1, \cdots, v_k) = \operatorname{span}(v_1, \cdots, v_{j-1}, v_{j+1}, \cdots, v_k)$.

Proof. Fix constants $a_1, \cdots, a_k \in \mathbb{F}$, not all zero, such that

$$a_1 \cdot v_1 + \cdots + a_k \cdot v_k = 0.$$

Let $j \coloneqq \max\{j \in \{1, \cdots, k\} \mid a_j \neq 0\}$. Then, $a_{j+1} = \cdots = a_k = 0$. We therefore have

$$v_j = \left(-\frac{a_1}{a_j}\right) \cdot v_1 + \cdots + \left(-\frac{a_{j-1}}{a_j}\right) \cdot v_{j-1} \in \operatorname{span}(a_1, \cdots, a_{j-1}). \tag{1}$$

Observe that by definition, $\operatorname{span}(v_1, \cdots, v_k) \supseteq \operatorname{span}(v_1, \cdots, v_{j-1}, v_{j+1}, \cdots, v_k)$. Now suppose $u \in \operatorname{span}(v_1, \cdots, v_k)$. Fix constants $b_1, \cdots, b_k$ such that

$$u = b_1 \cdot v_1 + \cdots + b_k \cdot v_k.$$

Then, substituting with Equation 1, we have

$$u = \left(b_1 - \frac{a_1}{a_j}\right) \cdot v_1 + \cdots + \left(b_{j-1} - \frac{a_{j-1}}{a_j}\right) \cdot v_{j-1} + b_{j+1} \cdot v_{j+1} + \cdots + b_k \cdot v_k \in \operatorname{span}(v_1, \cdots, v_{j-1}, v_{j+1}, \cdots, v_k).$$

Hence $\operatorname{span}(v_1, \cdots, v_k) \subseteq \operatorname{span}(v_1, \cdots, v_{j-1}, v_{j+1}, \cdots, v_k)$, and thus $\operatorname{span}(v_1, \cdots, v_k) = \operatorname{span}(v_1, \cdots, v_{j-1}, v_{j+1}, \cdots, v_k)$ □

The following result is crucial to building towards the concept of "dimension": the length of a linearly independent list is at most the length of a spanning list of a vector space.

Proposition 3.8. Suppose V is a vector space over $\mathbb{F}$. Let $u_1, \cdots, u_m$ be a linearly independent list of vectors in V and $v_1, \cdots, v_n$ a spanning list of V. Then, $m \leq n$.

Proof. Consider the following process, repeating step i for $i = 1, \cdots, m$. Let $k_s^1 = s$ for $s = 1, \cdots, n$. Let ℓ be the list $v_{k_1^1}, \cdots, v_{k_n^1}$.

Step *i*. Note that ℓ is a length-n spanning list of V of the form $u_1, \cdots, u_{i-1}, v_{k_1^i}, \cdots, v_{k_{n-i+1}^i}$. Let ℓ' be the list $u_1, \cdots, u_i, v_{k_1^i}, \cdots, v_{k_{n-i+1}^i}$ (i.e., inserting u_i to the i-th place in ℓ), which must be linearly dependent as u_i is a linear combination of ℓ. By Lemma 3.7, fix constants $a_1, \cdots, a_{n+1}$, not all zero, such that

$$a_1 \cdot u_1 + \cdots + a_i \cdot u_i + a_{i+1} \cdot v_{k_1^i} + \cdots + a_{n+1} \cdot v_{k_{n-i+1}^i} = 0. \tag{2}$$

Let $j \coloneqq \max\{j \in \{1, \cdots, n+1\} \mid a_j \neq 0\}$.

Suppose for the sake of contradiction that $j \leq i$. Then, $a_{j+1} = \cdots = a_i = \cdots = a_{n+1} = 0$. Hence, with $a_1, \cdots, a_j$ not all zero (in particular $a_j \neq 0$),

$$a_1 \cdot u_1 + \cdots + a_j \cdot u_j = 0,$$

which is impossible since $u_1, \cdots, u_j$, as a sub-list of $u_1, \cdots, u_m$, must be linearly independent.

Thus $j > i$, $\underline{\text{and thus } i+1 \leq j \leq n+1 \Rightarrow i \leq n \text{ due to the existence of such a } j}$. Note that a_j is multiplied with $v_{k^i_{j-i}}$. Then by the second conclusion in Lemma 3.7, $\operatorname{span}(u_1, \cdots, u_i, v_{k^i_1}, \cdots, v_{k^i_{n-i+1}}) = \operatorname{span}(u_1, \cdots, u_i, v_{k^i_1}, \cdots, v_{k^i_{j-i-1}}, v_{k^i_{j-i+1}}, \cdots, v_{k^i_{n-i+1}}) = V$. Let ℓ now be the length-n spanning list $u_1, \cdots, u_i, v_{k^i_1}, \cdots, v_{k^i_{j-i-1}}, v_{k^i_{j-i+1}}, \cdots, v_{k^i_{n-i+1}}$, and let k^{i+1}_s (where $s = 1, \cdots, n-i$) be the corresponding constants for vectors in ℓ from the list $v_1, \cdots, v_n$ by setting

$$k^{i+1}_1 = k^i_1, \cdots, k^{i+1}_{j-i-1} = k^i_{j-i-1}, k^{i+1}_j = k^i_{j-i+1}, \cdots, k^{i+1}_{n-i} = k^i_{n-i+1}.$$

Clearly, step 1 is feasible from the setup of ℓ and k^1_s for $s = 1, \cdots, n$. It has now been shown that each step i can be performed for $i = 1, \cdots, m$. Hence, the underlined predicate in i holds true in particular for $i = m$. Therefore, $m \leq n$. □

This proof is pretty crazy and I still feel a bit icky... But we can always prove this by supposing the contrary and setting up a linear system:

Proof. Suppose for the sake of contradiction that $m > n$. Define a "tall" matrix $A \in \mathbb{R}^{m \times n}$ of coefficients, each of whose rows contains at least one non-zero entry, such that

$$u_1 = A_{11} \cdot v_1 + \cdots + A_{1n} \cdot v_n,$$
$$\vdots$$
$$u_m = A_{m1} \cdot v_1 + \cdots + A_{mn} \cdot v_n.$$

Let $\tilde{A} \in \mathbb{R}^{m \times n}$ be a row echelon form of A. Then, (because $\tilde{A}$ is strictly "tall,") the $(n+1)$-th to the m-th rows of $\tilde{A}$ are all zero rows. In particular, because the m-th row of $\tilde{A}$ is the zero row, $u_m + v = 0$ for some $v \in \operatorname{span}(u_1, \cdots, u_{m-1})$, which would violate the given condition that $u_1, \cdots, u_m$ is an independent list. □

Another useful result: a subspace of a finite dimensional vector space is also finite dimensional.

Proposition 3.9. Let V be a finite dimensional vector space and $U \subseteq V$ a subspace. Then, U is finite dimensional.

Proof. Let $v_1, \cdots, v_n$ be a spanning list of V. Consider the following process for $i = 1, 2, \cdots$, setting ℓ to be the empty list of vectors in V.

Step *i*. Note that ℓ is a length-$(i-1)$ independent list of vectors in U. If $\operatorname{span} \ell = U$, then end the process. Otherwise, choose an arbitrary $u \in U \setminus \operatorname{span} \ell$ and append u to the end of ℓ. By the contrapositive of Lemma 3.7, ℓ is now a length-i linearly independent list of vectors in U.

It is obvious that the process will terminate after m steps where $m \leq n$, for otherwise Proposition 3.8 would be violated. Thus, the length-m list ℓ is independent and spans U. □

We now have the machinery to develop the concept of basis, and see how it gives rise to the notion of dimensionality.

Definition 3.10. Let V be a vector space over $\mathbb{F}$. A list of vectors in V is said to be a basis of V if the list is linearly independent and spans V.

The uniqueness of representation, a concept that we've already seen when dealing with direct sums of subspaces, also turns out to work here.

Proposition 3.11. Let V be a vector space over $\mathbb{F}$. A list of vectors $v_1, \cdots, v_k$ in V is a basis of V if and only if every $v \in V$ admits a unique choice of scalars $(a_1, \cdots, a_k) \in \mathbb{F}^k$ such that

$$v = a_1 \cdot v_1 + \cdots + a_k \cdot v_k.$$

Proof. **"If" direction.** Suppose any vector in V admits a unique linear combination of $v_1, \cdots, v_k$. In particular, 0 is a unique linear combination of $v_1, \cdots, v_k$, which must be $a_1 = \cdots = a_k = 0$, and hence the list is linearly independent. Thus, $v_1, \cdots, v_k$ is a basis of V.

"Only if" direction. Suppose instead that $v_1, \cdots, v_k$ is a basis of V, and fix scalars $(a_1, \cdots, a_k), (b_1, \cdots, b_k) \in \mathbb{F}^k$ such that

$$a_1 \cdot v_1 + \cdots + a_k \cdot v_k = b_1 \cdot v_1 + \cdots + b_k \cdot v_k.$$

Rearranging the terms,

$$(a_1 - b_1) \cdot v_1 + \cdots + (a_k - b_k) \cdot v_k = 0.$$

Because $v_1, \cdots, v_k$ is linearly independent, $a_1 - b_1 = \cdots = a_k - b_k = 0$. That is, $a_1 = b_1, \cdots, a_k = b_k$. The proof is now complete. □

The big idea behind the concept of a basis is that it's the only intersection of independent lists and spanning list, and this gives rise to the two following statements.

Proposition 3.12. Every spanning list of a vector space admits a sub-list that is a basis.

Proof. Suppose $\ell := v_1, \cdots, v_n$ is a spanning list. Consider the following process for $i = 1, \cdots, n$.

Step *i*. Note that v_i is an element of the spanning list ℓ. If <u>it is a linear combination of all vectors in ℓ previous to v_i</u>, then remove v_i from the list. Note that v_{i+1} remains an element of the list ℓ, and ℓ continues to span V by Lemma 3.7.

After the process is finished, ℓ is a linearly independent and spanning list. □

Consequently, every finite dimensional vector space, which already has a spanning list, must have a basis.

Proposition 3.13. Let V be a finite dimensional vector space. Then, V has a basis.

Proof. Let $v_1, \cdots, v_k$ be a spanning list of V. Then, by Proposition 3.12, it may be reduced to a sub-list that is a basis. □

The proof of the second statement directly uses the first.

Proposition 3.14. Every linearly independent list of a vector space admits a super-list that is a basis.

Proof. Let V be a vector space over $\mathbb{F}$. Suppose $u_1, \cdots, u_m$ is an independent list of vectors in V and let $v_1, \cdots, v_n$ be a spanning list of V. Then the joined list $u_1, \cdots, u_m, v_1, \cdots, v_n$ is also a spanning list of V. We now apply the process as described in the proof of Proposition 3.12, which gives a basis of V. Note that in the process, no u_j is ever removed for any $j \in \{1, \cdots, m\}$, for otherwise the list $u_1, \cdots, u_{j-1}$ would be linearly dependent, which is impossible. Hence the resultant list is a super-list of $u_1, \cdots, u_m$. □

To tie into the previous concept of direct sums of subspaces, we have the following:

Proposition 3.15. Let V be a finite dimensional vector space over $\mathbb{F}$ and suppose U is a subspace of V. Then, there exists a subspace W of V such that $U \oplus W = V$.

Proof. Let $u_1, \cdots, u_m$ be a basis of U, which we extend by Proposition 3.14 to $u_1, \cdots, u_m, w_1, \cdots, w_n$, a basis of W. Note that no w_i is a linear combination of $u_1, \cdots, u_m, w_1, \cdots, w_{i-1}$, so in particular it is not in U, the span of $u_1, \cdots, u_m$.

Let $W := \text{span}(w_1, \cdots, w_n)$. Now suppose for the sake of contradiction that there exists $(a_1, \cdots, a_m) \in \mathbb{F}^m \backslash \{0\}$ and $(b_1, \cdots, b_n) \in \mathbb{F}^n \backslash \{0\}$ such that

$$a_1 \cdot u_1 + \cdots + a_m \cdot u_m = b_1 \cdot w_1 + \cdots + b_n \cdot w_n \in U \cap W \backslash \{0\} \subseteq V.$$

Then, this particular vector admits two representations as linear combinations of the basis $u_1, \cdots, u_m, w_1, \cdots, w_n$, a contradiction. Hence, $U \cap W = \{0\}$, and thus the sum $U + W$ is direct.

Let $v \in V$ be an arbitrary vector, which admits a representation $v = c_1 \cdot u_1 + \cdots + c_m \cdot u_m + d_1 \cdot w_1 + \cdots + d_n \cdot w_n = (c_1 \cdot u_1 + \cdots + c_m \cdot u_m) + (d_1 \cdot w_1 + \cdots + d_n \cdot w_n) \in U \oplus W$. Therefore, $U \oplus W = V$. □

Now, a clever use of Proposition 3.8 tell us that all bases of a vector space have the same length, which becomes the foundation for the notion of dimensionality.

Proposition 3.16. Let V be a finite dimensional vector space over $\mathbb{F}$. Then, all bases of V have the same length.

Proof. Let ℓ_1, ℓ_2 be two spanning lists of V. Then, by Proposition 3.8, the length of ℓ_1 is at most that of ℓ_2. By symmetry, the length of ℓ_2 is at most that of ℓ_1. Hence ℓ_1 and ℓ_2 have the same length. □

Now the definition:

Definition 3.17. For each vector space V we associate a unique value, called the dimension of V and denoted as $\dim V$, in $\mathbb{Z}_{\geq 0} \cup \{+\infty\}$. If V is finite dimensional, let $\dim V \in \mathbb{Z}_{\geq 0}$ be the length of any basis of V. Otherwise, let $\dim V \coloneqq +\infty$.

Obviously, all subspaces are "smaller" than the space it is in, measured in terms of dimension. These super abstract statements are actually not hard to prove—we've finished a lot of those with some of the statements above.

Proposition 3.18. Suppose V is a finite-dimensional vector space and $U \subseteq V$ is a subspace. Then, $\dim U \leq \dim V$.

Proof. Note that U must be finite dimensional following Proposition 3.9. Now fix bases $u_1, \cdots, u_m$ of U and $v_1, \cdots, v_n$ of V. In particular $u_1, \cdots, u_m$ is linearly independent and $v_1, \cdots, v_n$ spans V. Hence, by Proposition 3.8, $m \leq n$. Therefore, by Proposition 3.16, $\dim U \leq \dim V$. □

The concept of dimension, which relies on Proposition 3.8, tells us this kind of "squeezing" happening: any independent list is no longer than a basis, and any spanning list is no shorter than a basis. Then, it's intuitive that an independent list long enough is a basis, and that a spanning list short enough is also a basis. The proof follows cleveryly from Propositions 3.12 and 3.14.

Proposition 3.19. Suppose V is a finite dimensional vector space and let $n \coloneqq \dim V$. Then, any independent list $v_1, \cdots, v_n$ is a basis of V.

Proof. Note that the list $v_1, \cdots, v_n$ may be extended to a super-list by Proposition 3.14. Suppose m new vectors are appended. Then, $n + m = \dim V$, so $m = \dim V - \dim V = 0$. Hence, the resultant list is $v_1, \cdots, v_n$, which is a basis. □

Proposition 3.20. Suppose V is a finite dimensional vector space and let $n \coloneqq \dim V$. Then, any spanning list $v_1, \cdots, v_n$ is a basis of V.

Proof. Note that the list $v_1, \cdots, v_n$ may be reduced to a sub-list by Proposition 3.12. Suppose m vectors are removed. Then, $n - m = \dim V$, so $m = \dim V - \dim V = 0$. Hence, the resultant list is $v_1, \cdots, v_n$, which is a basis. □

The concept of dimension is actually so well defined that it has an inclusion-exclusion property. We take for granted that for any subspaces U, W of a vector space V, the intersection $U \cap W$ is also a subspace, which obviously satisfies the closure properties.

Proposition 3.21. Suppose V is a finite-dimensional vector space and U, W are subspaces of V. Then, $\dim(U + W) = \dim U + \dim W - \dim(U \cap W)$.

Proof. Let V be an $\mathbb{F}$-vector space. Let $v_1, \cdots, v_k$ be a basis of $U \cap W$, which may be extended to a basis $v_1, \cdots, v_k, u_1, \cdots, u_m$ of U and a basis $v_1, \cdots, v_k, w_1, \cdots, w_n$ of W. Then, $\dim(U \cap W) = k$, $\dim U = k + m$, and $\dim W = k + n$. It suffices to show that $v_1, \cdots, v_k, u_1, \cdots, u_m, w_1, \cdots, w_n$ is a basis of $U + W$.

We first show linear independence. Consider constants $a_1, \cdots, a_k, b_1, \cdots, b_m, c_1, \cdots, c_n \in \mathbb{F}$ such that

$$a_1 \cdot v_1 + \cdots + a_k \cdot v_k + b_1 \cdot u_1 + \cdots + b_m \cdot u_m + c_1 \cdot w_1 + \cdots + c_n \cdot w_n = 0;$$

that is,

$$W \ni c_1 \cdot w_1 + \cdots + c_n \cdot w_n = (-a_1) \cdot v_1 + \cdots + (-a_k) \cdot v_k + (-b_1) \cdot u_1 + \cdots + (-b_m) \cdot u_m \in U,$$

so $c_1 \cdot w_1 + \cdots + c_n \cdot w_n \in U \cap W$, and thus admits a representation

$$c_1 \cdot w_1 + \cdots + c_n \cdot w_n = d_1 \cdot v_1 + \cdots + d_k \cdot v_k$$

for some constants $d_1, \cdots, d_k \in \mathbb{F}$; that is,

$$(-d_1) \cdot v_1 + \cdots + (-d_k) \cdot v_k + c_1 \cdot w_1 + \cdots + c_n \cdot w_n = 0.$$

Because $v_1, \cdots, v_k, w_1, \cdots, w_n$ is a basis and hence independent, we have $d_1 = \cdots = d_k = c_1 = \cdots = c_n = 0$.

Therefore,

$$(-a_1) \cdot v_1 + \cdots + (-a_k) \cdot v_k + (-b_1) \cdot u_1 + \cdots + (-b_m) \cdot u_m = 0,$$

and thus $a_1 = \cdots = a_k = b_1 = \cdots = b_m = 0$ since $v_1, \cdots, v_k, u_1, \cdots, u_m$ is a basis and hence independent.

Lastly, for any $v \in U + W$, there exists $u \in U$ and $w \in W$ such that $v = u + w$, where u and w each is a linear combination of the corresponding bases, whose sum is also representable as a linear combination of $v_1, \cdots, v_k, u_1, \cdots, u_m, w_1, \cdots, w_n$. The proof is complete. $\square$

This inclusion-exclusion doesn't work for more than 2 subspaces! This counterexample reminds us of the statement that the sum of three subspaces doesn't need to be direct even if their intersection is $\{0\}$!

However, when the sum of subspaces is distinct, we can add the dimensions up:

Corollary 3.22. Let V be a vector space and suppose $U_1, \cdots, U_n$ are finite-dimensional subspaces of V whose sum is direct. Then,

$$\dim(U_1 \oplus \cdots \oplus U_n) = \dim U_1 + \cdots + \dim U_n.$$

Proof. Note that

$$\begin{aligned}
\dim(U_1 \oplus \cdots \oplus U_n) &= \dim(U_1 + (U_2 \oplus \cdots \oplus U_n)) \\
&= \dim U_1 + \dim(U_2 \oplus \cdots \oplus U_n) \\
&= \dim U_1 + \dim U_2 + \dim(U_3 \oplus \cdots \oplus U_n) \\
&\vdots \\
&= \dim U_1 + \cdots + \dim U_n.
\end{aligned}$$

The proof is complete. $\square$

4 Linear Maps

4.1 Vector Space of Linear Maps

A linear map is just, well, a map that's linear.

Definition 4.1. A linear map T is a function from an $\mathbb{F}$-vector space V to an $\mathbb{F}$-vector space W that respects:

- $\forall u, v \in V, T(u+v) = T(u) + T(v)$;
- $\forall c \in \mathbb{F}, \forall v \in V, T(c \cdot v) = c \cdot T(v)$.

We denote the collection of all linear maps from V to W with $\mathcal{L}(V, W) = \hom(V, W)$.

Just like matrices, all we need are each column vectors—if the vector space is finite dimensional.

Proposition 4.2. Suppose $v_1, \cdots, v_n$ is a basis of V, and $w_1, \cdots, w_n \in W$. Then, there exists a unique $T \in \mathcal{L}(V, W)$ such that for any $i = 1, \cdots, n$, $T(v_i) = w_i$.

Proof. **Existence.** Let

$$T(c_1 \cdot v_1 + \cdots + c_n \cdot v_n) = c_1 \cdot w_1 + \cdots + c_n \cdot w_n,$$

which is a linear map and indeed satisfies $\forall i \in \{1, \cdots, n\}, T(v_i) = w_i$.

Uniqueness. Let $T(v_i) = w_i$ for $i = 1, \cdots, n$ for an arbitrary linear map $T \in \mathcal{L}(V, W)$. Then,

$$T(c_1 \cdot v_1 + \cdots + c_n \cdot v_n) = c_1 \cdot T(v_1) + \cdots + c_n \cdot T(v_n) = c_1 \cdot w_1 + \cdots + c_n + w_n,$$

hence T is unique. □

To get to the results we have for matrices, we should definitely define addition, scalar multiplication. This makes $\mathcal{L}(V, W)$ a vector space too.

Definition 4.3. Suppose $T, T_1, T_2 \in \mathcal{L}(V, W)$ and $c \in \mathbb{F}$, where V, W are both $\mathbb{F}$-vector spaces. Then, we define $(T_1 + T_2) \in \mathcal{L}(V, W)$ by setting

$$(T_1 + T_2)(v) = T_1(v) + T_2(v).$$

Similarly, we define $(c \cdot T) \in \mathcal{L}(V, W)$ by setting

$$(c \cdot T)(v) = c \cdot T(v).$$

Whenever possible, we define the product of two linear maps as their composition, which is also a linear map.

Proposition 4.4. Let V, W be vector spaces over $\mathbb{F}$. Then, $\mathcal{L}(V, W)$ is a vector space over $\mathbb{F}$.

Proof. It is obvious that:

- The 0 map is the additive identity;
- Additive commutativity follows from the additive commutativity of vectors;
- Additive associativity follows from the additive associativity of vectors;
- $(-1) \cdot T$ is the additive inverse of T;
- $1_{\mathbb{F}} \cdot T = T$ indeed;
- $((a \cdot b) \cdot T)(v) = (a \cdot b) \cdot T(v) = abT(v)$ equals $(a \cdot (b \cdot T))(v) = a \cdot (b \cdot T)(v) = abT(v)$;
- The two distributive properties follow from those of V and W.

Thus, $\mathcal{L}(V, W)$ is an $\mathbb{F}$-vector space. □

It wouldn't be surprising that $\dim \mathcal{L}(V, W) = (\dim V) \cdot (\dim W)$: you need that many numbers to fill the matrix for that map!

Proposition 4.5. Let V, W be finite-dimensional vector spaces over $\mathbb{F}$. Then, $\dim \mathcal{L}(V, W) = (\dim V) \cdot (\dim W)$.

The proof is essentially the same: fix a basis $v_1, \cdots, v_n$ of V and a basis $w_1, \cdots, w_m$ of W. Create the set of matrices, each of which has only one entry with 1: define $\{T_{i,j}\}_{(i,j)\in\{1,\cdots,n\}\times\{1,\cdots,m\}}$ by $T_{i,j}(v_k) := \delta_{k,i} w_j$. It's pretty boring here on out to show linear independence and the spanning property.

4.2 The Null Space and the Range

Definition 4.6. Let $T \in \mathcal{L}(V, W)$. The null space of T is the subspace

$$\operatorname{null} T = \{v \in V \mid T(v) = 0\}.$$

Let's check that it is a subspace: if $T(u) = T(v) = 0$ for some $u, v \in V$, then $(-v) \in V$ as well. So, $T(u) - T(-v) = T(u+v) = 0$. If $T(v) = 0$ for some $v \in V$, then $c \cdot T(v) = T(c \cdot v) = 0$ as well.

If a linear map T is injective, then in particular the pre-image of $\{0\}$ is the singleton (i.e., a set with exactly one element) is $\{0\}$. The null space gives this nice way of making this idea more general and precise:

Proposition 4.7. Suppose $T \in \mathcal{L}(V, W)$. Then, T is injective if and only if $\operatorname{null} T = \{0\}$.

Proof. **"If" direction.** Suppose $\operatorname{null} T = \{0\}$ and let $u, v \in V$ be such that $T(u) = T(v)$. Then, $T(u) - T(v) = T(u - v) = 0$, so $u - v \in \operatorname{null} T$, and hence $u - v = 0$. Therefore, $u = v$, and hence T is injective.

"Only if" direction. Suppose instead that $\operatorname{null} T \neq \{0\}$. Because $\operatorname{null} T$ is a subspace, it must contain a non-zero vector v. Then, $T(v) = T(0) = 0$, sop T is not injective. $\square$

I think I didn't have to resort to the contrapositive, but whatever works works.

To get to the fundamental theorem of linear maps, we need to establish that the range also has the structure of a vector space.

Proposition 4.8. Let $T \in \mathcal{L}(V, W)$. Then, $\operatorname{range} T$ is a subspace of W.

Proof. Let $\mathbb{F}$ be the field associated with V and W. Let $w_1, w_2 \in \operatorname{range} T$ and fix $u, v \in V$ such that $T(u) = w_1$ and $T(v) = w_2$. Then, $w_1 + w_2 = T(u) + T(v) = T(u + v) \in \operatorname{range} T$.

Now let $w \in \operatorname{range} T$ and $c \in \mathbb{F}$. Fix some $v \in V$ such that $T(v) = w$. Then, $c \cdot w = c \cdot T(v) = T(c \cdot v) \in \operatorname{range} T$. $\square$

And this is what all this machinery has been building up to:

Theorem 4.9 (Fundamental Theorem of Linear Maps)**.** *Suppose $T \in \mathcal{L}(V, W)$ where V is finite dimensional. Then,* $\dim \operatorname{null} T + \dim \operatorname{range} T = \dim V$.

Proof. Let $\mathbb{F}$ be the field associated with V and W. Because V is finite dimensional, so is its subspace $\operatorname{null} T$. Fix a basis $u_1, \cdots, u_m$ of $\operatorname{null} T$, so $\dim \operatorname{null} T = m$. Extend this list to a basis $u_1, \cdots, u_m, v_1, \cdots, v_n$ of V by Proposition 3.14, so $\dim V = m + n$.

It suffices now to show that $T(v_1), \cdots, T(v_n)$ is a basis of $\operatorname{range} T$. Let $w \in \operatorname{range} T$ and fix $v \in V$ such that $T(v) = w$. Fix constants $c_1, \cdots, c_n \in \mathbb{F}$ such that $v = c_1 \cdot v_1 + \cdots + c_n \cdot v_n$. Then, $T(v) = c_1 \cdot T(v_1) + \cdots + c_n \cdot T(v_n) \in \operatorname{span}(v_1, \cdots, v_n)$. Now let constants $c_1, \cdots, c_n \in \mathbb{F}$ instead be such that $c_1 \cdot T(v_1) + \cdots + c_n \cdot T(v_n) = 0$. Then, $T(c_1 \cdot v_1 + \cdots + c_n \cdot v_n) = 0$, so $(c_1 \cdot v_1 + \cdots + c_n \cdot v_n) \in \operatorname{null} T$ and thus may be expressed as

$$c_1 \cdot v_1 + \cdots + c_n \cdot v_n = d_1 \cdot u_1 + \cdots + d_m \cdot u_m$$

for some constants $d_1, \cdots, d_m \in \mathbb{F}$. Therefore,

$$(-d_1) \cdot u_1 + \cdots + (-d_m) \cdot u_m + c_1 \cdot v_1 + \cdots + c_n \cdot v_n = 0.$$

Because $u_1, \cdots, u_m, v_1, \cdots, v_n$ is a basis, it is linearly independent. Therefore, in particular, $c_1 = \cdots = c_n = 0$. The proof is now complete. □

Here are some relevant corollaries.

Corollary 4.10. Let $T \in \mathcal{L}(V, W)$ where $\dim V < \dim W < +\infty$. Then, T cannot be surjective.

Proof. It follows from Theorem 4.9 that

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

The final inequality remains strict. Therefore, $\operatorname{range} T \neq W$, and thus T is not surjective. □

Corollary 4.11. Let $T \in \mathcal{L}(V, W)$ where $\dim W < \dim V < +\infty$. Then, T cannot be injective.

Proof. It follows from Theorem 4.9 that

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

so $\operatorname{null} T \neq \{0\}$. Therefore, by Proposition 4.7, T is not injective. □

4.3 Matrices

Definition 4.12. Suppose $\mathbb{F}$ is a field and let $m, n \in \mathbb{N}_+$. An m-by-n $\mathbb{F}$-matrix, denoted as $A \in \mathbb{F}^{m,n}$, is a function $A\colon \{1, \cdots, m\} \times \{1, \cdots, n\} \to \mathbb{F}$, where $A_{i,j}$, called the (i, j)-th entry of A, is defined as $A(i, j)$ for $(i, j) \in \{1, \cdots, m\} \times \{1, \cdots, n\}$.

Pointwise addition and scalar multiplication of matrices as functions then make $\mathbb{F}^{m,n}$ a vector space.

We can see that $\mathbb{F}^{m,n}$ can be flattened to $\mathbb{F}^{mn}$, so their structures as vector spaces are identical. Indeed, they are isomorphic.

Definition 4.13. Let $T \in \mathcal{L}(V, W)$, where V and W are both finite dimensional $\mathbb{F}$-vector spaces. Suppose $v_1, \cdots, v_n$ is a basis of V and $w_1, \cdots, w_m$ a basis of W. Define the matrix of T under $v_1, \cdots, v_n$ and $w_1, \cdots, w_m$, denoted as $\mathcal{M}(T)$, by

$$\forall i \in \{1, \cdots, m\}, \qquad T(v_i) = \mathcal{M}(T)_{1,i} \cdot w_1 + \cdots + \mathcal{M}(T)_{m,i} \cdot w_m.$$

It's not entirely trivial that the matrix of a sum is the sum of the respective matrices! And the same goes for the scalar multiplications.

Proposition 4.14. Let $S, T \in \mathcal{L}(V, W)$ and $\lambda \in \mathbb{F}$, where V and W are both finite dimensional $\mathbb{F}$-vector spaces. Suppose $v_1, \cdots, v_n$ is a basis of V and $w_1, \cdots, w_m$ a basis of W. Then, $\mathcal{M}(S + T) = \mathcal{M}(S) + \mathcal{M}(T)$ and $\mathcal{M}(\lambda \cdot T) = \lambda \cdot \mathcal{M}(T)$.

Proof. Let $i \in \{1, \cdots, n\}$ be arbitrary. By definition, we have

$$\begin{cases} S(v_i) = \mathcal{M}(S)_{1,i} \cdot w_1 + \cdots + \mathcal{M}(S)_{m,i} \cdot w_m, \\ T(v_i) = \mathcal{M}(T)_{1,i} \cdot w_1 + \cdots + \mathcal{M}(T)_{m,i} \cdot w_m. \end{cases}$$

Then, adding the two,

$$\begin{aligned} S(v_i) + T(v_i) = (S+T)(v_i) &= (\mathcal{M}(S+T)_{1,i}) \cdot w_1 + \cdots + (\mathcal{M}(S+T)_{m,i}) \cdot w_m \\ &= (\mathcal{M}(S)_{1,i} + \mathcal{M}(T)_{1,i}) \cdot w_1 + \cdots + (\mathcal{M}(S)_{m,i} + \mathcal{M}(T)_{m,i}) \cdot w_m \end{aligned}$$

hence $\mathcal{M}(S+T) = \mathcal{M}(S) + \mathcal{M}(T)$.

Similarly,

$$\begin{aligned} \lambda \cdot T(v_i) = (\lambda \cdot T)(v_i) &= \mathcal{M}(\lambda \cdot T)(v_i)_{1,i} \cdot w_1 + \cdots + \mathcal{M}(\lambda \cdot T)(v_i)_{m,i} \cdot w_m \\ &= \lambda \cdot (\mathcal{M}(T)_{1,i} \cdot w_1 + \cdots + \mathcal{M}(T)_{m,i} \cdot w_m) \\ &= (\lambda \cdot \mathcal{M}(T)_{1,i}) \cdot w_1 + \cdots + (\lambda \cdot \mathcal{M}(T)_{m,i}) \cdot w_m, \end{aligned}$$

hence $\mathcal{M}(\lambda \cdot T) = \lambda \cdot \mathcal{M}(T)$. □

The matrix multiplication, while complicated-looking, is defined in that way for a reason:

Proposition 4.15. There exists a unique operation $\cdot : \mathbb{F}^{p,n} \times \mathbb{F}^{n,m} \to \mathbb{F}^{p,m}$ such that for any $\mathbb{F}$-vector spaces U, V and W with bases $u_1, \cdots, u_m$, $v_1, \cdots, v_n$, and $w_1, \cdots, w_p$ respectively, if $S \in \mathcal{L}(U,V)$ and $T \in \mathcal{L}(V,W)$, then $\mathcal{M}(TS) = \mathcal{M}(T) \cdot \mathcal{M}(S)$.

Proof. Let $i \in \{1, \cdots, m\}$ be arbitrary. Then,

$$\begin{aligned} &(TS)(u_i) \\ &= T(S(u_i)) \\ &= T(\mathcal{M}(S)_{1,i} \cdot v_1 + \cdots + \mathcal{M}(S)_{n,i} \cdot v_n) \\ &= \mathcal{M}(S)_{1,i} \cdot T(v_1) + \cdots + \mathcal{M}(S)_{n,i} \cdot T(v_n) \\ &= \mathcal{M}(S)_{1,i} \cdot (\mathcal{M}(T)_{1,1} \cdot w_1 + \cdots + \mathcal{M}(T)_{p,1} \cdot w_p) + \cdots + \mathcal{M}(S)_{n,i} \cdot (\mathcal{M}(T)_{1,n} \cdot w_1 + \cdots + \mathcal{M}(T)_{p,n} \cdot w_p) \\ &= (\mathcal{M}(T)_{1,1} \cdot \mathcal{M}(S)_{1,i} + \cdots + \mathcal{M}(T)_{1,n} \cdot \mathcal{M}(S)_{n,i}) \cdot w_1 + \cdots + (\mathcal{M}(T)_{p,1} \cdot \mathcal{M}(S)_{1,i} + \cdots + \mathcal{M}(T)_{p,n} \cdot \mathcal{M}(S)_{n,i}) \cdot w_p. \end{aligned}$$

Indeed, the definition $\mathcal{M}(TS)_{i,j} = \mathcal{M}(T)_{i,1} \cdot \mathcal{M}(S)_{1,j} + \cdots + \mathcal{M}(T)_{i,n} \cdot \mathcal{M}(S)_{n,j}$ is valid in the sense that $\mathcal{M}(TS) \in \mathbb{F}^{p,m}$. This is then the unique way to define matrix multiplication such that the given condition holds. □

4.4 Inverse Map

In general, a function $f : X \to Y$ has an inverse $f^{-1} : Y \to X$ iff f is bijective. For a linear map, we see that the inverse works out to be linear too!

Proposition 4.16. Let V, W be $\mathbb{F}$-vector spaces. If a linear map $T \in \mathcal{L}(V,W)$ is invertible, then the inverse $T^{-1} \in \mathcal{L}(W,V)$ is also linear.

Proof. Let I denote the identity map $\lambda w.\ w$ on W. Then, for any $c \in \mathbb{F}$ and $w_1, w_2 \in W$,

$$T(T^{-1}(c \cdot w_1 + w_2)) = I(c \cdot w_1 + w_2) = c \cdot I(w_1) + I(w_2) = c \cdot T(T^{-1}(w_1)) + T(T^{-1}(w_2)) = T(c \cdot T^{-1}(w_1) + T^{-1}(w_2)).$$

Because T is injective in particular, $T^{-1}(c \cdot w_1 + w_2) = c \cdot T^{-1}(w_1) + T^{-1}(w_2)$. □

Definition 4.17. Two $\mathbb{F}$-vector spaces V and W are said to be isomorphic, denoted as $V \cong W$, if there exists an invertible linear map $T \in \mathcal{L}(V,W)$. In this case, such a T is said to be an isomorphism from V to W.

Isomorphisms between vector spaces are really boring. In fact, it doesn't tell us much else than the dimension. More precisely:

Proposition 4.18. Suppose V and W are finite-dimensional $\mathbb{F}$-vector spaces. Then, $V \cong W$ if and only if $\dim V = \dim W$.

Proof. **"If" direction.** Fix bases $v_1, \cdots, v_n$ and $w_1, \cdots, w_n$ of V and W. Define $T \in \mathcal{L}(V, W)$ by

$$\forall i \in \{1, \cdots, n\}, \quad T(v_i) = w_i$$

through linear extension (Proposition 4.2). Then, $\operatorname{range} T = \operatorname{span}(w_1, \cdots, w_n) = W$, so T is surjective. Now suppose constants $c_1, \cdots, c_n \in \mathbb{F}$ are such that $T(c_1 \cdot v_1 + \cdots + c_n \cdot v_n) = c_1 \cdot w_1 + \cdots + c_n \cdot w_n = 0$. Because $w_1, \cdots, w_n$ is a linearly independent list, $c_1 = \cdots = c_n = 0$. Hence, $\operatorname{null} T = \{0\}$, and thus T is injective by Proposition 4.7.

"Only if" direction. Now suppose instead that $T \in \mathcal{L}(V, W)$ is invertible. Then, $\operatorname{null} T = \{0\}$ by Proposition 4.7. Because T is surjective, $\dim W = \dim \operatorname{range} T = \dim V - \dim \operatorname{null} T = \dim V$. □

A perhaps useful fact is given as follows.

Proposition 4.19. Let V, W be $\mathbb{F}$-vector spaces that share a finite dimension. Then, for any $T \in \mathcal{L}(V, W)$, the following are equivalent:

- T is invertible;
- T is injective;
- T is surjective.

Proof. **(1) ⇒ (2).** Suppose T is invertible. Then, it is injective in particular.

(2) ⇒ (3). Now suppose T is injective; that is, $\operatorname{null} T = \{0\}$ by Proposition 4.7. Then, $\dim \operatorname{range} T = \dim V - \dim \operatorname{null} T = \dim V$. Since $\operatorname{range} T \subseteq W$, it must be that $\operatorname{range} T = W$, and hence T is surjective.

(3) ⇒ (1). Finally, let T be surjective; that is, $\operatorname{range} T = W$. Then, $\dim \operatorname{null} T = \dim V - \dim \operatorname{range} T = \dim V - \dim W = 0$, so T is injective. Hence, T is bijective, and thus invertible. □

Corollary 4.20. Let V be a finite-dimensional vector space. Then, for any $T \in \mathcal{L}(V)$, the following are equivalent:

- T is invertible;
- T is injective;
- T is surjective.

4.5 The Product Space

Prototypical Example. $\mathbb{R} \times x\mathbb{R} \times x^2\mathbb{R} = \mathcal{P}_2[x \in \mathbb{R}]$.

Given any two $\mathbb{F}$-vector spaces, there is a natural way to define a new vector space on the Cartesian product with pointwise operations. A not-so-interesting but useful example is the $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. Of course, an n-fold Cartesian product is more general.

Definition 4.21. Let $V_1, \cdots, V_k$ be vector spaces. The (direct) product of $V_1, \cdots, V_k$, denoted with the regular Cartesian product as $V_1 \times \cdots \times V_k$, is the k-fold Cartesian product endowed with pointwise addition and pointwise scalar multiplication.

What does this space look like? We can view this from the perspective of bases. We'll restrict our discussions to the product of two vector spaces, but they can be readily generalized to more spaces.

Proposition 4.22. Suppose V, W are $\mathbb{F}$-vector spaces with bases $v_1, \cdots, v_n$ and $w_1, \cdots, w_m$ respectively. Then, $(v_1, 0), \cdots, (v_n, 0), (0, w_1), \cdots, (0, w_m)$ is a basis of $V \times W$.

Proof. We first show the spanning property. Let $(v, w) \in V \times W$ be arbitrary. Then, fix constants $c_1, \cdots c_n$ and $d_1, \cdots, d_m$ in

$\mathbb{F}$ such that $v = c_1 \cdot v_1 + \cdots + c_n \cdot v_n$ and $w = d_1 \cdot w_1 + \cdots + d_m \cdot w_m$. Thus,

$$\begin{aligned}(v, w) &= (c_1 \cdot v_1 + \cdots + c_n \cdot v_n, d_1 \cdot w_1 + \cdots + d_m \cdot w_m) \\ &= (c_1 \cdot v_1 + \cdots + c_n \cdot v_n, 0) + (0, d_1 \cdot w_1 + \cdots + d_m \cdot w_m) \\ &= c_1 \cdot (v_1, 0) + \cdots + c_n \cdot (v_n, 0) + d_1 \cdot (0, w_1) + \cdots + d_m \cdot (0, w_m).\end{aligned}$$

We now show linear independence. Suppose now that for constants $c_1, \cdots c_n$ and $d_1, \cdots, d_m$ in $\mathbb{F}$ we have

$$c_1 \cdot (v_1, 0) + \cdots + c_n \cdot (v_n, 0) + d_1 \cdot (0, w_1) + \cdots + d_m \cdot (0, w_m) = (v, w) = 0 = (0, 0),$$

where by the same logic we have defined $v := c_1 \cdot v_1 + \cdots + c_n \cdot v_n$ and $w := d_1 \cdot w_1 + \cdots + d_m \cdot w_m$. Hence, $v = w = 0$, and the linear independence of each basis guarantees $c_1 = \cdots = c_n = d_1 = \cdots = d_m = 0$. ◻

Corollary 4.23. Let V, W be finite-dimensional $\mathbb{F}$-vector spaces. Then, $\dim V \times W = \dim V + \dim W$.

Wait we add the dimensions? Isn't this supposed to be a product? It turns out there is *another* sense in which the direct product acts like a sum; namely:

Proposition 4.24. Let V, W be finite-dimensional $\mathbb{F}$-vector spaces. Then, $V \times W = V \times \{0\} \oplus \{0\} \times W$.

The proof is omitted. These products are really boring and don't give rise to new, interesting structures beside those of $\mathbb{F}^n$.

4.6 The Quotient Space

We don't have any sort of division in a vector space, but it turns out there is a way to "quotient" an entire space, at least in the sense of the dimensions working out to satisfy

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

which is kind of the reverse of $\dim V \times W = \dim V + \dim W$.

We now make this idea more precise.

Definition 4.25. Let V be a vector space and $U \subseteq V$ a subspace. Define $\underline{V/U} := \{v + U \mid v \in V\}$, where for any $v \in V$ we define $\underline{v + U} := \{v + u \mid u \in U\}$.

Intuitively, $v + U$ is just the resulting shape of moving the shape U by the vector v.

We'll put off the vector space business (i.e., defining $+$ and $\cdot$ on V/U) because there's actually a non-trivial issue that we haven't seen much of before: namely, one $\tilde{v} \in V/U$ can have two distinct representations $v_1 + U = v_2 + U$ (with $v_1 \neq v_2$). In fact, in many cases, each $\tilde{v} \in V/U$ has infinitely many representations.

For example, take V as the xy-plane $\mathbb{R}^2$ and $U = \{(x, y) \mid x = y \in \mathbb{R}\}$ the diagonal line. Then, $(0, 1) + U$ is the line $y = x + 1$, naturally. But $(1, 2)+U = \{(1, 2)+(x, x) \mid x \in \mathbb{R}\} = \{(x+1, x+2) \mid (x+1) \in \mathbb{R}\} = \{(x', y') \mid y' = x'+1 \in \mathbb{R}\} = (0, 1)+U$.

Fortunately, we have a very potent piece of machinery that'll help us through these weeds.

Lemma 4.26. Suppose V is a vector space with $U \subseteq V$ a subspace. Then, for any $v_1, v_2 \in V$, the following are equivalent.

- $v_1 + U = v_2 + U$;
- $(v_1 + U) \cap (v_2 + U) \neq \varnothing$.
- $v_1 - v_2 \in U$;

Proof. ($1 \implies 2$) Because $v_1 + 0 \in v_1 + U$, we know $v_1 + U = v_2 + U$ is non-empty. Thus, the intersection is non-empty.

($2 \implies 3$) Now fix a particular $v_1 + u_1 = v_2 + u_2 \in (v_1 + U) \cap (v_2 + U)$. Then, $v_1 - v_2 = (-1) \cdot u_1 + 1 \cdot u_2 \in U$.

($3 \implies 1$) Suppose $v_1 - v_2 \in U$, so $v_2 - v_1 \in U$ as well. Then,

$$v_1 + U = \{v_1 + u \mid u \in U\} = \{v_1 + \underbrace{(v_2 - v_1) + u'}_{=u} \mid u' \in U\} = \{v_2 + u' \mid u' \in U\} = v_2 + U.$$

The proof is finished. □

Now that we have the machinery, we'll first spell out what we want before justifying it makes sense.

Proposition 4.27. Let V be a vector space and $U \subseteq V$ a subspace. Then, V/U is made into a vector space with the following operations:

$$\begin{aligned}(v_1 + U) + (v_2 + U) &\coloneqq (v_1 + v_2) + U,\\ c \cdot (v + U) &\coloneqq (c \cdot v) + U.\end{aligned}$$

The reason I call this a Proposition is because there is a need to justify that the definitions are consistent; that is, different representations $v + U = v' + U$ of the same object gives the same answer/output. We use the prime here to keep track of representations since we already used the symbols v_1 and v_2 for addition.

Proof. Suppose $v_1 + U = v_1' + U$ and $v_2 + U = v_2' + U$, so $v_1 - v_1', v_2 - v_2' \in U$. Then,

$$(v_1 + U) + (v_2 + U) = (v_1 + v_2) + U = (v_1 + v_2 + (v_1' - v_1) + (v_2' - v_2)) + U = (v_1' + v_2') + U = (v_1' + U) + (v_2' + U).$$

Similarly, suppose $v + U = v' + U$, so $c \cdot (v' - v) \in U$. By the same logic,

$$c \cdot (v + U) = (c \cdot v) + U = (c \cdot v + c \cdot (v' - v)) + U = (c \cdot v') + U = c \cdot (v' + U).$$

The proof is complete. □

To get to the desired equality about the dimensions, we'll obviously use a map and apply the fundamental theorem of linear maps. A natural choice is:

Definition 4.28. Suppose V is a vector space. To each subspace $U \subseteq V$ is associated a map $\pi \colon V \longrightarrow V/U$, called the quotient map, defined as

$$\pi(v) \coloneqq v + U.$$

Proposition 4.29. Suppose V is a finite-dimensional $\mathbb{F}$-vector space and $U \subseteq V$ is a subspace. Then, $\dim V/U = \dim V - \dim U$.

Well, we already have $\dim V$ and $\dim \operatorname{range} \pi = \dim V/U$. All we need to show is $\dim U = \dim \operatorname{null} \pi$.

Proof. Fix a basis $u_1, \cdots, u_m$ of U and extend by Proposition 3.14 to a basis $u_1, \cdots, u_m, v_1, \cdots, v_n$ of V, so $\dim U = m$ and $\dim V = m + n$. Now,

$$\operatorname{null} \pi = \{v \in V \mid v + U = 0_{V/U}\} = \{v \in V \mid v \in U\} = U.$$

Thus, by Theorem 4.9, $\dim V/U = \dim \operatorname{range} \pi = \dim V - \dim \operatorname{null} \pi = \dim V - \dim U$. □

So far things are boring and static. To get more structure, we'll look at linear maps from or on V/U. In particular, we'll see how "quotient"-ing the null space gives rise to a very nice map.

Proposition 4.30. Suppose V, W are vector spaces. To each $T \in \mathcal{L}(V, W)$ we may associate a linear map $\tilde{T} \in \mathcal{L}(V/\operatorname{null} T, W)$ by

$$\tilde{T}(v + \operatorname{null} T) \coloneqq T(v).$$

As before, we'll need to show that this definition is consistent; that is, two representations of the same input give the same output.

Proof. Suppose $v + \operatorname{null} T = v' + \operatorname{null} T$; that is, $v - v' \in \operatorname{null} T$ (Lemma 4.26), so $-T(v - v') = 0$. Then, $\tilde{T}(v + \operatorname{null} T) = T(v) = T(v) - T(v - v') = T(v') = \tilde{T}(v' + \operatorname{null} T)$. $\square$

Like we've seen, it turns out the notation $v + U$ for representing an element in V/U, while not unique, plays nicely with linear stuff. Now, $\operatorname{null} T$ is precisely where $T(v) = 0$; i.e., where T fails to be injective. Now that we "quotient"-ed away the null space, it is reasonable to expect the resultant $\tilde{T}$ is injective. In fact, we have an even stronger result.

Proposition 4.31. Suppose V, W are $\mathbb{F}$-vector space and $T \in \mathcal{L}(V, W)$. Then, $\tilde{T}$ is a isomorphism from $V/\operatorname{null} T$ to $\operatorname{range} T$.

Proof. Surjectivity is obvious by construction. We now show injectivity. Suppose $\tilde{T}(v + \operatorname{null} T) = \tilde{T}(v' + \operatorname{null} T)$; that is, $T(v) = T(v')$. Thus, $T(v - v') = 0$, so $v - v' \in \operatorname{null} T$. Thus, by Lemma 4.26, $v + \operatorname{null} T = v' + \operatorname{null} T$. $\square$

4.7 The Dual Space

Prototypical Example. Row vectors.

We've seen $\mathbb{R}^n$, the set of n-tuples of real numbers. We can also think of them as column vectors, just a different way to write n-tuples. So what about $1 \times n$ row vectors?

Thinking more generally as to what these row vectors do, I claim it's more appropriate to think of them as **functions**. Consider for example, the 2×1 row vector $v' := \begin{pmatrix} 1 & 2 \end{pmatrix}$. I'm saying that v' is a function that takes column vectors to numbers by:

$$v' : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1 + 2x_2.$$

That is, $v'(v)$ is just the matrix multiplication $v' \cdot v$. Indeed, v' is a function that takes $\mathbb{R}^2$, the set of column vectors, to $\mathbb{R}$.

The reason we want to aim for this apparently weird and roundabout way of thinking about these row vectors is because functions—linear maps, actually—are readily applicable to more general discussions of vector spaces.

Definition 4.32. To each vector space V over $\mathbb{F}$ we associate its <u>dual space</u> $V' := \mathcal{L}(V, \mathbb{F})$.

We know the set of linear maps is a vector space, so every dual space is a vector space. One thing we can immediately say, then, is:

Proposition 4.33. Let V be a finite-dimensional vector space. Then, $\dim V' = \dim V$.

Proof. By Proposition 4.5, $\dim V' = \dim \mathcal{L}(V, \mathbb{F}) = (\dim V) \cdot 1 = \dim V$. $\square$

So now what? Because we defined the dual space in terms of functions, the abstract way to go, it gets harder to actually do concrete stuff/calculations like the example above. But all we need is a basis: that's where matrices/tables of numbers come from. We always want to keep in mind the "row vector" intuition as we develop more abstract concepts around the dual space. No one can just get this without a "picture."

Definition 4.34. Let V be a vector space. To each basis $v_1, \cdots, v_n$ of V we associate a length-n list of vectors $\phi_1, \cdots, \phi_n$ in V', called the <u>dual basis</u> of $v_1, \cdots, v_n$, which is defined by

$$\forall i \in \{1, \cdots, n\} \qquad \phi_i(v_j) := \delta_{i,j}.$$

Here, $\delta_{i,j}$, called the Kronecker delta, is 1 if $i = j$ and 0 otherwise. This definition is valid because all we need to define a linear map ϕ_i is to define what values it takes on the basis.

By the way, just because we call this list the dual basis doesn't mean it's actually a basis. We need to show that's true.

Proposition 4.35. Let V be a vector space and $v_1, \cdots, v_n$ a basis of V. Then, the dual basis $\phi_1, \cdots, \phi_n$ is a basis of V'.

Proof. We first show linear independence. Let $\mathbb{F}$ be the field associated with V. Suppose constants $c_1, \cdots, c_n \in \mathbb{F}$ are such that

$$c_1 \cdot \phi_1 + \cdots + c_n \cdot \phi_n = 0.$$

Then, for any e_i where $i = 1, \cdots, n$,

$$(c_1 \cdot \phi_1 + \cdots + c_n \cdot \phi_n)(e_i) = c_1 \cdot \phi_1(e_i) + \cdots + c_n \cdot \phi_n(e_i) = c_1\delta_{1,i} + \cdots + c_n\delta_{n,i} = c_i = 0.$$

Thus, $\phi_1, \cdots, \phi_n$ is a length-n, linearly independent list of vector in V'. Because $\dim V' = \dim V = n$, the dual basis is also a spanning list by Proposition 3.14, and thus a basis. □

What about linear maps? How do the idea of linear maps and that of the dual space interact? More specifically, we want some sort of a dual map.

Invoking the intuition of making a column vector a row vector by taking the transpose, we similarly want to define a dual map that acts like the transpose of a matrix. Because we use the abstract language of functions, it becomes hard to track things without making clear the domains and codomains. If $T \in \mathcal{L}(V, W)$ with $\dim V = n$ and $\dim W = m$ (i.e., an $m \times n$ matrix), we want a "transpose" T' (i.e., an "$n \times m$" matrix) that takes an m-dimensional object to an n-dimensional object. We will do this by taking $T' \in \mathcal{L}(W', V')$, instead of the perhaps more intuitive choice of $T' \in \mathcal{L}(W, V)$ as an actual $n \times m$ matrix, because the following definition is so much more natural:

Definition 4.36. Let V, W be vector spaces over the same field. To each linear map $T \in \mathcal{L}(V, W)$ we associate a unique <u>dual map</u> $T' : W' \to V'$, defined as

$$T'(\psi) \coloneqq \psi \circ T.$$

That's it, just the composition! This works because this $\psi \circ T$, which is a linear function in V', takes in a vector $v \in V$ and would output $(\psi \circ T)(v) = \psi(T(v))$. Now $T(v) \in W$, so we can shove it inside ψ just fine and get a number. As we will show, this natural operation preserves linearity just fine.

Note that the composition of two functions $f \circ g$ is linear in f when g is fixed. This is because $((cf_1 + f_2) \circ g)(x) = (cf_1 + f_2)(g(x)) = cf_1(g(x)) + f_2(g(x)) = (c \cdot (f_1 \circ g))(x) + (f_2 \circ g)(x) = (c \cdot (f_1 \circ g) + f_2 \circ g)(x)$. In general, $f \circ g$ may not be linear in g, but it does work out if f is also linear.

Proposition 4.37. Let V, W be vector spaces over the same field. For every $T \in \mathcal{L}(V, W)$, its dual map $T' : W' \to V'$ is a linear map.

Proof. It is obvious that

$$T'(c \cdot \psi_1 + \psi_2) = (c \cdot \psi_1 + \psi_2) \circ T = c \cdot \psi_1 \circ T + \psi_2 \circ T = c \cdot T'(\psi_1) + T'(\psi_2).$$

The proof is finished. □

The usual properties of the matrix transpose readily translates to the map duality, except the proofs... it gets hectic.

Proposition 4.38. Let $S \in \mathcal{L}(V, W)$ and $T, T_1, T_2 \in \mathcal{L}(U, V)$. Let $\mathbb{F}$ denote the field associated with U, V, W and suppose $c \in \mathbb{F}$. Then,

- $(c \cdot T_1 + T_2)' = c \cdot T_1' + T_2'$;
- $(ST)' = T'S'$.

While these results seem obvious, the proofs are surprisingly slippery. Like I said, the linearity of $f \circ g$ in g is not immediately obvious. For the following proof, always keep in mind what the object is: $\psi \in V'$ is a function from V to $\mathbb{F}$; the output $T'(\psi) \in U'$ is a function from U to $\mathbb{F}$. For proving the second point, we use the fact that $\circ$ is associative: $(f \circ g) \circ h = f \circ (g \circ h)$ naturally.

Proof. Firstly, for any $\psi \in V'$ and $u \in U$,

$$
\begin{aligned}
(c \cdot T_1 + T_2)'(\psi)(u) &= (\psi \circ (c \cdot T_1 + T_2))(u) && \text{(Definition of $'$)} \\
&= \psi((c \cdot T_1 + T_2)(u)) && \text{(Definition of $\circ$)} \\
&= \psi(c \cdot T_1(u) + T_2(u)) && \text{(Pointwise VS ops on $\mathcal{L}(U, V)$)} \\
&= c \cdot \psi(T_1(u)) + \psi(T_2(u)) && \text{(ψ is linear)} \\
&= c \cdot (\psi \circ T_1)(u) + (\psi \circ T_2)(u) && \text{(Definition of $\circ$)} \\
&= (c \cdot (\psi \circ T_1) + \psi \circ T_2)(u) && \text{(Pointwise VS ops on U')} \\
&= (c \cdot T_1'(\psi) + T_2'(\psi))(u) && \text{(Definition of $'$)} \\
&= (c \cdot T_1' + T_2')(\psi)(u) && \text{(Pointwise VS ops on $\mathcal{L}(V', U')$)}
\end{aligned}
$$

so $(c \cdot T_1 + T_2)' = c \cdot T_1' + T_2'$.

To show the second item, let $\omega \in W'$ be arbitrary. Then,

$$(ST)'(\omega) = \omega \circ (S \circ T) = (\omega \circ S) \circ T = S'(\omega) \circ T = T'(S'(\omega)) = (T' \circ S')(\omega),$$

so $(ST)' = T'S'$. □

[TODO: Annihilators]

5 Some Results on Polynomials

To develop the next big topic—eigen-stuff, we need some more machinery on polynomials. As we recall, we frequently used the (monic) "characteristic polynomial" $\chi_A(\lambda) = \det(A - \lambda \cdot I)$ for a square matrix A. When we need results related to existence, these results will be greatly helpful.

Recall that for $n \in \mathbb{Z}_{\geq 0}$, the polynomial vector spaces over a field $\mathbb{F}$ are defined as

$$
\begin{aligned}
\mathcal{P}_n(\mathbb{F}) &= \{x \mapsto a_0 + \cdots + a_n x^n \mid a_0, \cdots, a_n \in \mathbb{F}\}, \\
\mathcal{P}(\mathbb{F}) &= \bigcup_{n=0}^{\infty} P_n(\mathbb{F}).
\end{aligned}
$$

For any $n \in \mathbb{Z}_{\geq 0}$, the vectors $e_i \coloneqq x \mapsto x^i$ for $i = 0, \cdots, n$ form a basis of $\mathcal{P}_n(\mathbb{F})$, so $\dim \mathcal{P}_n(\mathbb{F}) = n + 1$.

For a polynomial $p \in \mathbb{P}(\mathbb{F})$ with $p(x) = a_0 + \cdots + a_n$, we have $\deg p = \max\{j \in \{0, \cdots, n\} \mid a_j \neq 0\}$ if $p \neq 0$. Otherwise, $\deg 0 = -\infty$. This definition, while weird, allows us to have:

$$
\begin{aligned}
\deg(p + q) &= \max\{p, q\}, \\
\deg(pq) &= \deg p + \deg q,
\end{aligned}
$$

where we have intuitively $-\infty + x = -\infty$ and $-\infty < x$ for any $x \in \mathbb{F}$. We omit the proof to stay focused on the linear algebra aspect.

Here is a really neat application of linear algebra: the degree and the dimension of the polynomial subspace has a nice link that allows us to prove existence and uniqueness with the bijectivity of a certain map.

Proposition 5.1 (Polynomial Division)**.** For any $p, s \in \mathcal{P}(\mathbb{F})$ with $s \neq 0$, there exists unique $q \in \mathcal{P}_{n-m}(\mathbb{F})$ and $r \in \mathcal{P}_{m-1}(\mathbb{F})$ such that $p = sq + r$.

The proof below might look a bit messy, but draw the parallel to regular integer division. If $n < m$, we have something similar to $2 \div 5 = 0$ with remainder 2, which is sort of trivial. Otherwise, we have a non-zero quotient, which is where linear algebra comes in.

Proof. Let $n = \deg p$ and $m = \deg s$ and suppose $q \in \mathcal{P}_{n-m}(\mathbb{F})$ and $r \in \mathcal{P}_{m-1}(\mathbb{F})$ are arbitrary.

Suppose $n < m$. If $q \neq 0$, then

$$p = sq + r \Rightarrow n = \deg(sq + r) = \max\{m + \underbrace{\deg q}_{\geq 0}, \underbrace{\deg r}_{\leq m-1}\} = m + \deg q \geq m,$$

a contradiction. Thus, $q = 0$, so $p = sq+r = 0+r$ implies $r = p$ necessarily. Indeed, $(q, r) = (0, p)$ uniquely satisfies $p = sq+r$.

Now suppose $n \geq m$. Similarly, $p = sq + r \Rightarrow n = m + \deg q$, so $\deg q = n - m$ and thus $q \in \mathcal{P}_{n-m}(\mathbb{F})$. Define a map $T\colon \mathcal{P}_{n-m}(\mathbb{F}) \times \mathcal{P}_{m-1}(\mathbb{F}) \to \mathcal{P}_n(\mathbb{F})$ by $T(q, r) \coloneqq sq + r$. Note that T is linear; i.e., $T \in (\mathcal{P}_{n-m}(\mathbb{F}) \times \mathcal{P}_{m-1}(\mathbb{F}), \mathcal{P}_n(\mathbb{F}))$: indeed,

$$(s \cdot (cq) + (cr)) + (s \cdot \tilde{q} + \tilde{r}) = s \cdot (cq + \tilde{q}) + (cr + \tilde{r}).$$

We claim that T is invertible. Note that

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

so it suffices to show that T is injective by Proposition 4.19. Suppose $T(q, r) = sq + r = 0$. Then, $q = 0$, for otherwise

$$\deg(sq + r) = \max\{m + \underbrace{\deg q}_{\geq 0}, \underbrace{\deg r}_{\leq m-1}\} \geq m \neq -\infty.$$

Further, $sq + r = s \cdot 0 + r = 0$, so $r = 0$ also. Then, $\operatorname{null} T = \{(0, 0)\} = \{0\}$, so T is injective and hence bijective as claimed. Thus, for every $p \in \mathcal{P}_n(\mathbb{F})$, there exists unique $q \in \mathcal{P}_{n-m}(\mathbb{F})$ and $r \in \mathcal{P}_{m-1}(\mathbb{F})$ as claimed. □

The most important result related to our discussions is the fundamental theorem of algebra, which we state below. We also omit the proof because it must rely on some sort of algebra-related results.

Theorem 5.2 (Fundamental Theorem of Algebra)**.** *Any non-constant complex polynomial $p \in \mathcal{P}(\mathcal{C})$ has at least one zero.*

Corollary 5.3 (Polynomial Decomposition)**.** For any non-constant complex polynomial $p \in \mathcal{P}(\mathbb{C})$ (that is, $\deg p \geq 1$), there are unique $c \in \mathbb{C}$ and $\{\lambda_1, \cdots, \lambda_n\} \subset \mathbb{C}$ such that $p(z) = c \cdot (z - \lambda_1) \cdots (z - \lambda_n)$, where $n = \deg p$.

Proof. Let $n \coloneqq \deg p \geq 1$ and fix constants $a_0, \cdots, a_n \in \mathbb{F}$ such that $p(z) = a_0 + \cdots + a_n x^n$. Then by Theorem 5.2, p has a root λ_1. Let $s(z) \coloneqq z - \lambda_1$, so $\deg s = 1$. By Proposition 5.1, fix $q \in \mathcal{P}_{n-1}(\mathbb{C})$ and $r \in \mathcal{P}_0(\mathbb{C})$ such that $p = sq + r$.

Suppose for the sake of contradiction that $r \neq 0$. Then, $p(\lambda_1) = 0 + r(\lambda_1) = 0$. But $r \in \mathcal{P}_0(\mathbb{C})$ is a constant polynomial, so $r = 0$, a contradiction. Therefore, $r = 0$, so $p(z) = (z - \lambda_1) \cdot q(z)$, where $\deg q = n - 1$. Now, set $p \longleftarrow q$ and repeat the procedure n times, for all of which we have $\deg n \geq 1$, to obtain zeroes $\lambda_1, \cdots, \lambda_n$. Now $\deg p \leq 0$, so p is a constant $c = p(0)$. Thus,

$$p(z) = c \cdot (z - \lambda_1) \cdots (z - \lambda_n).$$

The proof is complete. □

Note that the λ_i's can repeat; that is, $|\{\lambda_1, \cdots, \lambda_n\}| \leq n$ in general. For this reason, we might prefer the following form which has unique $c \in \mathbb{C}$ and unique $\lambda_1', \cdots, \lambda_m' \in \mathbb{C}$ with $\lambda_1' < \cdots < \lambda_m'$:

$$p(z) = c \cdot (z - \lambda_1')^{d_1} \cdots (z - \lambda_m')^{d_m},$$

where $d_1, \cdots, d_m \in \mathbb{N}_+$ and $d_1 + \cdots + d_m = n$—we obtain the m distinct roots with their algebraic multiplicity being $d_1, \cdots, d_m$ respectively.

6 Eigen-Stuff

6.1 Basic Definitions and Facts

When we studied the equation $Tv = \lambda v$ for matrices, we had an abundance of results related to eigenvalues and such. The equation, though, work also in general for any operators, which allows us to extend this discussion far beyond just matrices. The applications are everywhere: from solving differential equations to recurrences, they show up in all practical areas of science and engineering.

Before we go right in to eigenvalues, let's first look at what abstract properties a map should have to have an eigenvalue. We use the language of invariance:

Definition 6.1. Suppose V is a finite-dimensional vector space and $U \subseteq V$ is a subspace. Let $T \in \mathcal{L}(V)$. Then, U is said to be invariant under T if $T(U) \subseteq U$.

Because T is an operator from V to V, some structures get nicely preserved naturally. It's not hard to see that if $Tv = \lambda v$ for some $v \neq 0$, then $\operatorname{span}(v)$ is invariant under T. Conversely, if a 1-d subspace $U = \operatorname{span}(v)$ is invariant under T for some $v \neq 0$, then $T|_U \in \mathcal{L}(U)$ must be a scalar multiple, and thus correspond to an eigenvalue.

For some examples, $\{0\}$, V, $\operatorname{null} T$, and $\operatorname{range} T$ are all invariant under T always.

But for defining the eigenvalue, we will content ourselves with the usual formulation.

Definition 6.2. Suppose V is an $\mathbb{F}$-vector space and $T \in \mathcal{L}(V)$. A number $\lambda \in \mathbb{F}$ is said to be an eigenvalue of T if $T(v) = \lambda v$ for some $v \in V \backslash \{0\}$. Let $E(T) := \{\lambda \in \mathbb{F} \colon \exists v \in V \backslash \{0\}, T(v) = \lambda \cdot v\}$ be the set of all T's eigenvalues. If $\lambda \in E(T)$, then we define the associated eigenspace $E(T, \lambda) := \{v \in V \mid T(v) = \lambda \cdot v\}$, which is a subspace of V.

Now that we've defined eigenvalues, how do we find one? We might recall solving $\det(T - \lambda \cdot I) = 0$ for λ. While we do not have the determinant yet, we know that $\det A = 0$ iff A is not-invertible. In the language we are familiar with, it turns out injectivity suffices to capture this property. Note that invertibility and injectivity of operators **are not**, in general, equivalent in (possibly infinite-dimensional) vector spaces. While the determinant no longer exists for infinite operators (specifically, infinite matrices), the definitions of eigen-stuff for those operators remain valid, and are in fact really useful in a wide variety of settings like physics, differential equations, etc.

Proposition 6.3. Suppose V is an $\mathbb{F}$-vector space, $T \in \mathcal{L}(V)$, and $\lambda \in \mathbb{F}$. Then, $\lambda \in E(T)$ if and only if $T - \lambda \cdot I$ is not injective.

Proof. **"If" direction.** Suppose $T - \lambda \cdot I$ is not injective. Then, fix $v \in V \backslash \{0\}$ such that $(T - \lambda \cdot I)(v) = Tv - \lambda \cdot Iv = 0$, so $Tv = \lambda v$ for some $v \neq 0$ and hence $\lambda \in E(T)$.

"Only if" direction. Now suppose $\lambda \in E(T)$. By definition, fix $v \in V \backslash \{0\}$ such that $Tv = \lambda v$. Then, $(T - \lambda \cdot I)(v) = 0$ but $v \neq 0$, so $T - \lambda \cdot I$ has a non-trivial null space and thus is not injective. □

Logically, we introduce eigenvectors after eigenvalues, because an eigenvalue usually corresponds to many eigenvectors.

Definition 6.4. Suppose $T \in \mathcal{L}(V)$ and $\lambda \in E(T)$. A vector $v \in V \backslash \{0\}$ is said to be an eigenvector of T associated with λ, or a λ-eigenvector of T, if $Tv = \lambda v$.

For some examples:

- For the identity operator $I \in \mathcal{L}(V)$, we have $E(I) = \{1\}$ and any non-zero vector is an eigenvector;
- For the derivative operator $T \in \mathcal{L}(\mathcal{P}(\mathbb{F}))$ with $T(p) = p'$, we have $E(T) = \{0_{\mathbb{F}}\}$ and no eigenvectors exist.
- For the (counterclockwise) rotation-by-90-degrees operator $T \in \mathcal{L}(\mathbb{R}^2)$, $E(T) = \varnothing$ and consequently no eigenvectors

exist.

As we saw, there are many more eigenvectors than eigenvalues, so would they “collide?” That is, how do eigenvectors associated with different eigenvalues act? It turns out that they're nicely independent.

Proposition 6.5. Suppose $T \in \mathcal{L}(V)$. If $\lambda_1, \cdots, \lambda_n$ are distinct eigenvalues of T and $v_1, \cdots, v_n$ are corresponding eigenvectors respectively, then $v_1, \cdots, v_n$ is linearly independent.

This proof is far from trivial. The big idea is to use proof by contradiction and invoke the powerful linear dependence lemma.

Proof. Let $\mathbb{F}$ be the field associated with V. Suppose for contradiction that $v_1, \cdots, v_n$ is linearly dependent. By Lemma 3.7, fix the smallest $j \in \{1, \cdots, n\}$ such that $v_j \in \operatorname{span}(v_1, \cdots, v_{j-1})$. Note that $v_1, \cdots, v_{j-1}$ is linearly independent.

Fix constants $c_1, \cdots, c_{j-1} \in \mathbb{F}$ such that $v_j = c_1 \cdot v_1 + \cdots + c_{j-1} \cdot v_{j-1}$. Applying T to both sides,

$$\lambda_j \cdot v_j = c_1\lambda_1 \cdot v_1 + \cdots + c_{j-1}\lambda_{j-1} \cdot v_{j-1}.$$

If we alternatively multiply by λ_j on both sides,

$$\lambda_j \cdot v_j = c_1\lambda_j \cdot v_1 + \cdots + c_{j-1}\lambda_j \cdot v_{j-1}.$$

Subtracting the two yields

$$c_1(\lambda_1 - \lambda_j) \cdot v_1 + \cdots + c_{j-1}(\lambda_{j-1} - \lambda_j) \cdot v_{j-1} = 0.$$

The linear independence of $v_1, \cdots, v_{j-1}$ forces $c_1(\lambda_1 - \lambda_j) = \cdots = c_{j-1}(\lambda_{j-1} - \lambda j) = 0$. Because the eigenvalues are distinct, $\lambda_1 - \lambda_j, \cdots, \lambda_{j-1} - \lambda_j$ are all nonzero, so $c_1 = \cdots = c_{j-1} = 0$, which implies $v_j = c_1 \cdot v_1 + \cdots + c_{j-1} \cdot v_{j-1} = 0$, a contradiction since eigenvectors are non-zero. □

Corollary 6.6. Suppose V is finite-dimensional. Then, any $T \in \mathcal{L}(V)$ has at most $\dim V$ distinct eigenvalues.

Proof. If $T \in \mathcal{L}(V)$ has $m > \dim V$ distinct eigenvalues, then they correspond to m linearly independent eigenvectors. But Proposition 3.14, $m \leq \dim V$, a contradiction. □

Albeit not the perfect time, we will now introduce the quotient operator T/U on an invariant subspace U.

Proposition 6.7. Suppose $T \in \mathcal{L}(V)$. To each invariant subspace U under T, we may associate a quotient operator $T/U \in \mathcal{L}(V/U)$ by $(T/U)(v + U) := T(v) + U$.

To use this definition, we need $T(v) + U$ to make sense; that's why the codomain of T must also be in V, so T must be an operator. And if U is not invariant, this definition might not be consistent. By the way, we also need to show that T/U is linear.

Proof. Suppose $v_1, v_2 \in V$ are such that $v_1 + U = v_2 + U$; that is, $v_1 - v_2 \in U$. Then, $T(v_1) - T(v_2) = T(v_1 - v_2) \in U$ as well, so $T(v_1) + U = T(v_2) + U$. Therefore, the definition is consistent.

We now show linearity. Observe that $(T/U)(c \cdot (v_1 + U) + (v_2 + U)) = (T/U)((c \cdot v_1 + v_2) + U) = T(c \cdot v_1 + v_2) + U = (c \cdot T(v_1) + T(v_2)) + U = c \cdot (T(v_1) + U) + (T(v_2) + U) = c \cdot (T/U)(v_1) + (T/U)(v_2)$. Therefore, T is indeed linear. □

One main aspect as to why eigen-stuff are useful is because they work well with repeatedly applying operators: this is repeatedly multiplying by the eigenvalue.

Definition 6.8. For an operator $T \in \mathcal{L}(V)$ and a non-negative integer $n \in \mathbb{Z}_{\geq 0}$, define the power

$$T^n := \underbrace{T \circ \cdots \circ T}_{n \text{ times}}.$$

If $n = 0$, then $T^n := I$.

We also introduce a new piece of machinery: the polynomial of an operator.

Definition 6.9. Suppose V is an $\mathbb{F}$-vector space, $T \in \mathcal{L}(V)$, and $p \in \mathcal{P}(\mathbb{F})$. Fix $p(x) = a_0 + \cdots + a_n x^n$. Then, define

$$\underline{p(T)} := a_0 I + \cdots + a_n T^n \in \mathcal{L}(V).$$

This is where the previous results on polynomials come in handy: this will help us prove the existence of an eigenvalue for non-trivial, finite-dimensional complex vector spaces.

Theorem 6.10. *Suppose V is a finite-dimensional, non-zero complex vector space. Then any operator $T \in \mathcal{L}(V)$ has at least one eigenvalue.*

Proof. Fix $T \in \mathcal{L}(V)$. Define a map $f_T : \mathcal{P}(\mathbb{C}) \to \mathcal{L}(V)$ by $f_T(p) := p(T)$. It is not hard to see that f_T is a linear map from an infinite-dimensional vector space to a finite-dimensional vector space, which cannot be injective.[1]

Thus, fix a nontrivial (i.e. non-constant) $p \in \operatorname{null} f_T \backslash \{0\}$ such that $p(T) = 0$. By Corollary 5.3, fix $c, \lambda_1, \cdots, \lambda_m \in \mathbb{C}$ such that $p(z) = c \cdot (z - \lambda_1) \cdots (z - \lambda_m)$. It is not hard to see that $p(T) = c \cdot (T - \lambda_1 I) \cdots (T - \lambda_m I) \in \mathcal{L}(V)$.

Suppose for contradiction $T - \lambda_i \cdot I$ is injective for all $i \in \{1, \cdots, m\}$. Then, for any $v \in V \backslash \{0\}$,

$$(T - \lambda_1 I)(v), \quad (T - \lambda_2 I)(T - \lambda_1 I)(v), \quad \cdots, \quad (T - \lambda_m I) \cdots (T - \lambda_1 I)(v)$$

must all be non-zero, so $p(T)(v) \neq 0$, which contradicts $p(T) = 0$. Thus, T has at least one eigenvalue λ_i by Proposition 6.3; i.e., the i for which $T - \lambda_i \cdot I$ is not injective. $\square$

Note that both premises are important:

- In a real vector space, even if it's finite-dimensional, operators may not have eigenvalues. For example, the $90°$ counterclockwise rotation matrix in $\mathbb{R}^2$ has no eigenvalue (think geometrically).
- In an infinite-dimensional vector space, even if it's complex, operators may not have eigenvalues. For example, take $V = \mathcal{P}(\mathbb{R})$ and let $T \in \mathcal{L}(V)$ be defined by $T(p)(x) := x \cdot p'(x)$. Then, T has no eigenvalues (why?).

A final note that relates to quantum mechanics: Solving Schrödinger's equation is essentially solving the eigenvalue problem

$$\hat{H}|\psi\rangle = E|\psi\rangle$$

with a given operator $\hat{H}$ (the Hamiltonian) to identify the eigenvalue E (the energy) and the eigenvector $|\psi\rangle$ (the stationary state). The operator is on a space of (wave) functions, which is an infinite-dimensional complex vector space.

In infinite dimensions, operators may have a finite number of eigenvalues, countably many eigenvalues, or even uncountably many eigenvalues. It is even possible that $E(T) = \mathbb{F}$.

6.2 Upper-Triangular Matrices

From MATH 220, we know that for upper-triangular matrices, we can just read off its diagonals for eigenvalues. This result is by no means a coincidence, and we'll explore this further.

Definition 6.11. A square matrix $A \in \mathbb{F}^{n,n}$ is said to be upper-triangular if for any $i \in \{1, \cdots, n\}$ and $j \in \{i + 1, \cdots, n\}$, $A_{i,j} = 0$.

[1] To see this, first note that an injective map takes independent vectors to independent vectors (the converse is true even for non-injective maps). Then, a sufficiently long independent list from the infinite-dimensional domain must be mapped to an independent list that is prohibitively long in the finite-dimensional codomain, thus producing a contradiction.

For an upper-triangular matrix A, its i-th column has zeros after the i-th entry, so $T(v_i)$ is a combination of $v_1, \cdots, v_i$ where $A = \mathcal{M}(T)$. Extending this intuition a bit more, we have this result:

Proposition 6.12. Let V be a finite-dimensional, non-zero complex vector space and $v_1, \cdots, v_n$ a basis of V. Then, the following are equivalent.

- $\mathcal{M}(T)$ is upper-triangular;
- $T(v_j) \in \operatorname{span}(v_1, \cdots, v_j)$ for any $j \in \{1, \cdots, n\}$;
- $\operatorname{span}(v_1, \cdots, v_j)$ is invariant under T for any $j \in \{1, \cdots, n\}$.

We will omit the proof; it is quite straightforward.

The reason upper-triangular matrices are of interest to us is because of its intimate relationship with eigenvalues: in a sense, almost every operator has an upper-triangular matrix for some basis.

Theorem 6.13. *Let V be a finite-dimensional, non-zero complex vector space and $T \in \mathcal{L}(V)$. Then, there exists some basis of V under which $\mathcal{M}(T)$ is upper-triangular.*

Proof. Let $n = \dim V$ be the dimension of V, upon which we perform induction.

Base case. Suppose $n = 1$. Then, any $n \times n$ matrix is trivially upper-triangular.

Inductive case. Suppose now that any operator on any $(n-1)$-dimensional complex vector space admits an upper-triangular matrix representation, where $n \in \mathbb{Z}_{\geq 2}$. Let V be an n-dimensional complex vector space and $T \in \mathcal{L}(V)$. By Theorem 6.10, fix an eigenvalue $\lambda_1 \in \mathbb{C}$ and an associated eigenvector $v_1 \in V \backslash \{0\}$. Let $U := \operatorname{span}(v_1)$, which is a 1-dimensional subspace; hence, V/U has dimension $\dim V - \dim U = n-1$, so $T/U \in \mathcal{L}(V/U)$ admits an upper-triangular matrix representation $\mathcal{M}(T/U)$ under some basis $v_2+U, \cdots, v_n+U$. By Proposition 6.12, we have for an arbitrary $i \in \{2, \cdots, n\}$ that $(T/U)(v_i + U) = T(v_i) + U \in \operatorname{span}(v_2 + U, \cdots, v_i + U)$; that is, for some $c_2, \cdots, c_i \in \mathbb{C}$,

$$T(v_i) + U = c_2 \cdot (v_2 + U) + \cdots + c_i \cdot (v_i + U) = (c_2 \cdot v_2 + \cdots + c_i \cdot v_i) + U,$$

which implies $T(v_i) - (c_2 \cdot v_2 + \cdots + c_i \cdot v_i) \in U$ by Lemma 4.26. Because U is a 1-dimensional subspace, we must have $T(v_i) - (c_2 \cdot v_2 + \cdots + c_i \cdot v_i) = \lambda \cdot v_1$ for some $\lambda \in \mathbb{C}$, so $T(v_i) \in \operatorname{span}(v_1, \cdots, v_i)$. At the same time, $T(v_1) = \lambda_1 \cdot v_1 \in \operatorname{span}(v_1)$.

If $v_2, \cdots, v_n$ are linearly dependent, then $v_2 + U, \cdots, v_n + U$ would also be linearly dependent. Taking the contrapositive, we conclude that $v_2, \cdots, v_n$ is linearly independent. If $v_1 \in \operatorname{span}(v_2, \cdots, v_n)$, then

$$0_{V/U} = v_1 + U \in \operatorname{span}(v_2 + U, \cdots, v_n + U),$$

a contradiction. Thus, by Lemma 3.7, $v_1, \cdots, v_n$ is a length-n linearly independent list in V, and as such must be a basis. Therefore, by Proposition 6.12, $\mathcal{M}(T)$ is upper-triangular with respect to the basis $v_1, \cdots, v_n$. □

If we have found an upper-triangular matrix representation of T, then a lot of the properties are evident. Of these, a useful fact is that T is invertible, or an isomorphism, if and only if all diagonals of the upper-triangular $\mathcal{M}(T)$ are non-zero.

Proposition 6.14. Let V be a vector space and $T \in \mathcal{L}(V)$. Suppose $\mathcal{M}(T)$ is upper-triangular with respect to a basis $v_1, \cdots, v_n$. Then, T is invertible if and only if $\mathcal{M}(T)_{i,i} \neq 0$ for all $i \in \{1, \cdots, n\}$.

Proof. Denote with $\lambda_i := \mathcal{M}(T)_{i,i}$ for $i \in \{1, \cdots, n\}$ the i-th diagonal entry.

"If" direction. Suppose $\lambda_i \neq 0$ always. By the upper-triangular representation, in particular, $T(v_1) = \lambda_1 \cdot v_1$, so $T(v_1/\lambda_1) = v_1 \in \operatorname{range} T$. More generally, for any $i \in \{1, \cdots, n\}$, $v_i \in \operatorname{range} T$.

To show this, we perform induction on $i = 1, \cdots, n$, where the base case has already been shown. Now suppose inductively that $v_1, \cdots, v_i \in \operatorname{range} T$. Observe that

$$T(v_i) = \mathcal{M}(T)_{1,1} \cdot v_1 + \cdots + \mathcal{M}(T)_{i-1,i-1} \cdot v_{i-1} + \lambda_i \cdot v_i.$$

Because each $v_1, \cdots, v_{i-1}$ is in $\operatorname{range} T$, we may fix constants for their respective linear combinations. Summing them, we have constants $c_1, \cdots, c_{i-1} \in \mathbb{F}$ such that

$$T(v_i) = c_1 \cdot T(v_1) + \cdots + c_{i-1} \cdot T(v_{i-1}) + \lambda i \cdot v_i,$$

so $T(v_i/\lambda_i - c_1/\lambda_i \cdot v_1 - \cdots - c_{i-1}/\lambda_i \cdot v_{i-1}) = v_i \in \operatorname{range} T$.

Thus, $\operatorname{range} T$ contains a basis of V so $\operatorname{range} T = V$. Then, by Corollary 4.20, T is invertible.

"Only if" direction. Suppose now $\lambda_j = 0$ for some $j \in \{1, \cdots, n\}$. Then, $T(v_j) \in \operatorname{span}(v_1, \cdots, v_{j-1})$, and for any previous $i \in \{1, \cdots, j-1\}$, we have $T(v_i) \in \operatorname{span}(v_1, \cdots, v_i) \subset \operatorname{span}(v_1, \cdots, v_{j-1})$. Therefore, we have a total of j vectors $T(v_1), \cdots, T(v_j)$ in a $(j-1)$-dimensional subspace $\operatorname{span}(v_1, \cdots, v_{j-1})$, so the vectors must be linearly dependent. An injective linear function maps linearly independent vectors to linearly independent vectors, so T cannot be injective. By Corollary 4.20, T is not invertible. □

Surprisingly, the very famous (and non-trivial) fact that the diagonal elements of an upper-triangular matrix are exactly its eigenvalues comes immediately after this result; the proof is so simple.

Corollary 6.15. Let V be a finite-dimensional vector space and suppose $T \in \mathcal{L}(V)$ has an upper-triangular matrix under some basis $v_1, \cdots, v_n$ of V. Then, every diagonal element is an eigenvalue of T; that is,

$$\forall i \in \{1, \cdots, n\}, \qquad \mathcal{M}(T)_{i,i} \in E(T).$$

Proof. Let $i \in \{1, \cdots, n\}$ be arbitrary and let $\lambda_i := \mathcal{M}(T)_{i,i}$. Then, naturally, $\mathcal{M}(T - \lambda_i \cdot I)_{i,i} = 0$, so by Proposition 6.14, T is not invertible—and hence not injective (Corollary 4.20). Thus, $\lambda_i \in E(T)$ by Proposition 6.3. □

6.3 Diagonal Matrices

What about diagonal matrices? This ties into the $A = P^{-1}\Lambda P$ decomposition for which Λ is diagonal, and we've already seen in MATH 220 how this can be used to compute high powers of A easily.

Note its form: I claim that surrounding Λ by P^{-1} and P is exactly a change of basis! If $A = \mathcal{M}(T)$ under some basis $e_1, \cdots, e_n$, then Λ is the matrix of T under the basis $P(e_1), \cdots, P(e_n)$. So, similar to upper-triangular matrices, we only need to see if the matrix of an operator *can* be diagonal under *some* basis.

We'll first lay out the definition.

Definition 6.16. A matrix $A \in \mathbb{F}^{n,n}$ is said to be diagonal if $\forall i, j \in \{1, \cdots, n\}$, $i \neq j \implies A_{i,j} = 0$. An operator $T \in \mathcal{L}(V)$ over a finite-dimensional vector space V is said to be diagonalizable if there exists a basis $v_1, \cdots, v_n$ under which $\mathcal{M}(T)$ is diagonal.

Unlike "upper-triangularizability," the diagonalizability of a matrix isn't guaranteed even in finite-dimensional, non-zero complex vector spaces. But before we justify this with an example, we'll need some potent machinery.

Proposition 6.17. Let V be a finite-dimensional vector space and let $\{\lambda_1, \cdots, \lambda_m\} = E(T)$. Then, T is diagonalizable if and only if there exist eigenvectors $v_1, \cdots, v_n$ of T that form a basis of V.

Proof. We opt for a direct proof in each direction.

"If" direction. Suppose $v_1, \cdots, v_n$ are eigenvectors of T associated with (possibly repeated) eigenvalues $\lambda_1, \cdots, \lambda_n$ that form a basis of V. Then, under this basis $v_1, \cdots, v_n$, we have by definition $T(v_i) = \lambda_i \cdot v_i$ for all $i \in \{1, \cdots, n\}$. Thus, the i-th column of $\mathcal{M}(T)$ is the tuple

$$(0, \cdots, 0, \underbrace{\lambda_i}_{\text{the } i\text{-th position}}, 0, \cdots, 0).$$

By definition, then, $\mathcal{M}(T)$ is diagonal, and thus T is diagonalizable.

"Only if" direction. Now suppose instead that T is diagonalizable. Fix a basis $v_1, \cdots, v_n$ such that $\mathcal{M}(T)$ is diagonal, whereby we may fix $\lambda_1, \cdots, \lambda_n \in \mathbb{F}$ such that $T(v_i) = \lambda_i \cdot v_i$ for each $i \in \{1, \cdots, n\}$. Because $v_1, \cdots, v_n$ is a basis, it must be a linearly independent list. In particular, no $v_i = 0$ for $i \in \{1, \cdots, n\}$. Thus, by definition, $v_1, \cdots, v_n$ are all eigenvectors of T, associated with $\lambda_1, \cdots, \lambda_n$ respectively. ◻

We have a very useful corollary:

Corollary 6.18. Let V be a finite-dimensional vector space and suppose $T \in \mathcal{L}(V)$ has $\dim V$ distinct eigenvalues. Then, T is diagonalizable.

Proof. Let $n := \dim V$ and suppose $\lambda_1, \cdots, \lambda_n$ are the n distinct eigenvalues of T. By definition, we may fix a λ_i-eigenvector $v_i \in V \backslash \{0\}$ for each $i \in \{1, \cdots, n\}$. By Proposition 6.5, $v_1, \cdots, v_n$ must be linearly independent. Thus, $v_1, \cdots, v_n$ is a length-n, linearly independent list of vectors, which must be a basis of V by Proposition 3.14. ◻

The converse is not true! Consider, for example, the matrix

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

which is by definition a diagonalizable operator on $\mathbb{R}^3$. This is also an upper-triangular matrix, so its eigenvalues are 1 and 0 (and no others). It doesn't have 3 distinct eigenvalues, but it is still diagonalizable.

When we do linear algebra, we want abstract structures beyond those on $\mathbb{R}^n$ uncovered in MATH 220. A crucial piece is the direct sum. We will first state an intuitively true but important lemma.

Lemma 6.19. Suppose $T \in \mathcal{L}(V)$. If $\lambda_1, \cdots, \lambda_k$ are distinct eigenvalues of T, then

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

is a direct sum.

Proof. Let $u_1 \in E(T, \lambda_1), \cdots, u_k \in E(T, \lambda_k)$ and suppose $u_1 + \cdots + u_k = 0$. Let $u_{i_1}, \cdots, u_{i_j}$ be *all* the j non-zero vectors among $u_1, \cdots, u_k$, so

$$1 \cdot u_{i_1} + \cdots + 1 \cdot u_{i_j} = 0,$$

where $u_{i_1}, \cdots, u_{i_j}$ must be linearly independent by Proposition 6.5. If $j > 0$, then this is impossible, because this equation would be a non-trivial linear combination that equals 0.

Therefore, $j = 0$; that is, $u_1 = \cdots = u_k = 0$. Therefore, by Proposition 2.11, the sum $E(T, \lambda_1) + \cdots + E(T, \lambda_k)$ is direct. ◻

Note that V doesn't have to be finite-dimensional here! So long as the eigenvalues are distinct, this results holds. This, then, allows us to tie in diagonalizability with the abstract structure of the direct sum.

Proposition 6.20. Let V be a finite-dimensional vector space and suppose $\lambda_1, \cdots, \lambda_k$ are all the (distinct) eigenvalues of $T \in \mathcal{L}(V)$. Then, T is diagonalizable if and only if $E(T, \lambda_1) \oplus \cdots \oplus E(T, \lambda_k) = V$.

Proof. Let $\mathbb{F}$ be the field associated with V.

"If" direction. Suppose $E(T,\lambda_1) \oplus \cdots \oplus E(T,\lambda_k) = V$. For each λ_i where $i \in \{1,\cdots,k\}$, we may fix a basis $v_1^{(i)},\cdots,v_{m_i}^{(i)}$ of $E(T,\lambda_i) \cong \mathbb{F}^{m_i}$. First, note that $m_1 + \cdots + m_k = n := \dim V$ by Proposition 3.22. Now, because the direct sum is V, $v_1^{(1)},\cdots,v_{m_1}^{(1)},\cdots,v_1^{(k)},\cdots,v_{m_k}^{(k)}$ must span V. Therefore, this list of eigenvectors is a basis of V by Proposition 3.12.

"Only if" direction. Now suppose instead that T is diagonalizable. Fix a basis $v_1,\cdots,v_n$ of V such that $\mathcal{M}(T)$ is diagonal, so by definition $T(v_i) = \ell_i \cdot v_i$ for each $i \in \{1,\cdots,n\}$, where $\ell_i := \mathcal{M}(T)_{i,i} \in E(T)$. Because a diagonal matrix is trivially upper-triangular, we conclude that $\{\ell_1,\cdots,\ell_n\} = \{\lambda_1,\cdots,\lambda_k\}$. Thus, for any $v \in V$ we may fix $c_1,\cdots,c_n \in V$ such that

$$v = c_1 \cdot v_1 + \cdots + c_n \cdot v_n \in E(T,c_1) + \cdots + E(T,c_n) = E(T,\lambda_1) \oplus \cdots \oplus E(T,\lambda_k).$$

The proof is complete. $\square$

Note that $k \neq n$ in general! For example, take a look again at

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

To continue the previous claim that diagonalizability cannot really be guaranteed for general cases, consider the operator $T \in \mathbb{C}^2$ defined by $T(w,z) := (z,0)$. We'll work it out in full detail:

Firstly, $E(T) = \{0\}$: let $\lambda \in \mathbb{C}$ be arbitrary and suppose $(w,z) \in \mathbb{C}^2 \backslash \{0\}$ is such that $T(w,z) = (z,0) = \lambda \cdot (w,z)$; that is, $z = \lambda w$ and $0 = \lambda z$. If $\lambda \neq 0$, then $0 = \lambda z$ implies $z = 0$, so $z = 0 = \lambda w$ and thus $w = 0$ as well, a contradiction. Now, 0 is indeed an eigenvalue, because $T(1,0) = (0,0) = 0 \cdot (1,0)$, where $(1,0) \in \mathbb{C}^2 \backslash \{0\}$. More specifically,

$$E(T,0) = \{(w,z) \in \mathbb{C}^2 \mid (z,0) = (0,0)\} = \{(w,0) \mid w \in \mathbb{C}\} = \operatorname{span}((1,0)).$$

Now, the sum of all eigenspaces is simply $E(T,0) = \operatorname{span}((1,0))$ itself, which is only 1-dimensional and cannot possibly equal $V = \mathbb{C}^2$. Thus, by Proposition 6.20, T is not diagonalizable.

7 Notes on the Determinant and the Trace

We'll first say that these concepts only make sense for finite-dimensional operators! That's why we've always introduced them in terms of square matrices.

7.1 The Determinant

Determinants are used to describe area- and volume-like stuff. For example, one of its biggest uses are in multivariable calculus, where the Jacobian (determinant) is used when we do change of variables, which describes how the infinitesimal area/volume changes for the given parametrization.

The determinant also keeps track of the "orientation" of the space. For example, $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ has determinant 1, but when you swap the columns, $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ has determinant -1. This does make sense intuitively: we basically flipped the space $\mathbb{F}^2$ with respect to the diagonal line $y = x$.

We'll start with the familiar formula from MATH 220, which in a 3×3 case has:

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \cdot \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \cdot \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \cdot \begin{vmatrix} d & e \\ g & h \end{vmatrix} = aei - afh - bdi + bfg + cdh - ceg.$$

Let's look at the first term in a bit more detail. The formula crosses out the row and the column where a is, and does this recursively. So in each term of the final expansion, we have elements from separate rows and columns multiplied together. There are, naturally, 3! such terms. But how do we account for the pluses and minuses?

Like we have noted above, this keeps track of the "orientation," or the flipping, of a space. While not immediately obvious, this turns out to be related to how we get the separate rows and columns. We'll first state the general Leibniz's formula, and go from there.

Definition 7.1. For a positive integer $n \in \mathbb{N}_+$, let S_n denotes the set (of size $n!$) of all permutations of the set $\{1, \cdots, n\}$; that is, the set of bijections from $\{1, \cdots, n\}$ to $\{1, \cdots, n\}$. We typically write σ_i instead of $\sigma(i)$, and we can write a permutation concisely as a tuple $(\sigma_1, \cdots, \sigma_n)$.

The sign of a permutation $\sigma \in S_n$, denoted as $\operatorname{sgn} \sigma$, is defined as $(-1)^k$, where k is the number of swaps to make to get from $(1, \cdots, n)$ to $(\sigma_1, \cdots, \sigma_n)$.

Some notes on the definition above:

- We can replace the bijection requirement with the equivalent condition that $\sigma_1, \cdots, \sigma_n$ is just $1, \cdots, n$ written in a different order; in other words, $\{1, \cdots, n\} = \{\sigma_1, \cdots, \sigma_n\}$.
- The k for a permutation is not unique, but they all give the same $(-1)^k$, which makes the definition consistent. For example, $(1, 3, 2) \in S_3$ can have $k = 1$, when we swap the 2 and the 3 in $(1, 2, 3)$ once, or $k = 3$, if we swap them three times. While not trivial, we skip the proof due to irrelevance.
- The sign can be alternatively defined as

  $$\operatorname{sgn} \sigma := \left|\{(i, j) \in \{1, \cdots, n\} \times \{1, \cdots, n\} \mid i < j, \sigma_i > \sigma_j\}\right|,$$

  the number of "inversion" pairs in σ.

Definition 7.2. Let $A \in \mathbb{F}^{n,n}$. The determinant of A, denoted as $\det A$ or $|A|$, is defined as

$$\sum_{\sigma \in S_n} \left(\operatorname{sgn} \sigma \cdot \prod_{i=1}^{n} A_{\sigma_i, i} \right).$$

Yikes... That looks terrible, but the good thing about this definition (instead of, say the one I gave above) is that we have explicitly each expanded term, which makes it easier to prove stuff about the determinant.

This definition is not as contrived as it may seem: in fact, this is the only definition of a determinant that satisfies the following properties (see Wikipedia):

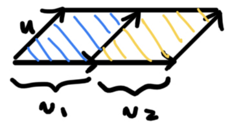
- The determinant is multilinear; that is, for any $i \in \{1, \cdots, n\}$, fixing all the other columns $1, \cdots, i-1, i+1, \cdots, n$, the determinant as the function of column i (from $\mathbb{F}^n$ to $\mathbb{F}$) is linear;
- The determinant is alternating; that is, switching any two columns will reverse the sign of the determinant;
- The determinant of any identity matrix I_n is 1.

The last property makes sense, because the "hyper-volume" of a hyper-cube is obviously $1 \cdots 1 = 1$. But how do we make sense of the other properties? We can, in fact, turn to geometric intuition.

For multilinearity, think about two parallelograms with the same height.

Fixing u constant, clearly you should add up the areas. That is,

$$\begin{vmatrix} u_1 & (v_1 + v_2)_1 \\ u_2 & (v_1 + v_2)_2 \end{vmatrix} = \begin{vmatrix} u_1 & (v_1)_1 \\ u_2 & (v_1)_2 \end{vmatrix} + \begin{vmatrix} u_1 & (v_2)_1 \\ u_2 & (v_2)_2 \end{vmatrix}.$$



The same goes for scalar multiples (just think of $v_2 = v_1$, in which case that would be doubling the second column).

What about the alternating property? Well, swapping any two column "flips" the space: consider the matrices

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

By swapping the two columns, we are effectively flipping the orientation of the resulting shape, so it makes sense that the determinant gets multiplied by -1.

Here are some other properties of the determinant:

- (Homogeneity) $\det(c \cdot A) = c^n \cdot \det A$ (each column get multiplied by c, so what does multilinearity say?);
- (Alternating) If any two columns of A are equal, then $\det A = 0$ (what does the alternating property say if you swap the two identical columns?);
- Adding (a scalar multiple of) any column to any *other* column doesn't change the determinant (think about multilinearity and the alternating property);
- If any column is zero, then $\det A = 0$ (what happens if you add any other column to this column?).

These are just some immediate consequences of the definition we have. But that's not all that the determinant has to offer.

Arguably the most amazing property of the determinant is its multiplicativity.

Theorem 7.3. *Let $A, B \in \mathbb{F}^{n,n}$. Then,*

$$\det(AB) = \det A \cdot \det B.$$

I will have to omit the proof here: an elementary proof exists by checking the definition, but this conveniently overlooks the inherent structure that the determinant has. Evan Chen's *Napkin* has an elegant explanation for why this is: it's so much more than this innocent formula seemingly entails!

This is so powerful that we have an extremely strong corollary.

Corollary 7.4. Let $A \in \mathbb{F}^{n,n}$. When viewed as a linear operator $A \in \mathcal{L}(\mathbb{F}^n)$, A is invertible if and only if

$$\det A \neq 0.$$

We'll prove only the "only if" direction. For the "if" direction, this can be shown by proving Cramer's rule—which in my humble opinion is also not that theoretically interesting.

Proof. **"Only if" direction.** Suppose A is invertible. Suppose for the sake of contradiction that $\det A = 0$. Then, $AA^{-1} = I$, so $\det(A)\det(A^{-1}) = 0$. But this must equal $\det(AA^{-1}) = \det I = 1$, a contradiction. □

The multiplicativity of the determinant means that the determinant of any matrix of a linear operator is the same. This allows us to define the determinant of such an operator.

Definition 7.5. Let V be a finite-dimensional vector space and $T \in \mathcal{L}(V)$. The <u>determinant of T</u> is defined as the determinant of the matrix of T under any basis.

It is not immediately obvious that this definition is consistent: different matrices should, in general, have different determinants; so how can we assert they're always the same?

We need to introduce the concept of change-of-basis matrices.

Definition 7.6. Suppose V is a finite-dimensional vector space over $\mathbb{F}$. Let $\mathcal{V} = \{v_1, \cdots, v_n\}$ and $\mathcal{W} = \{w_1, \cdots, w_n\}$ be two bases of V. The change-of-basis matrix from $\mathcal{V}$ to $\mathcal{W}$, denoted as $C_{\mathcal{W}\leftarrow\mathcal{V}} \in \mathbb{F}^{n,n}$, is defined as

$$C_{\mathcal{W}\leftarrow\mathcal{V}} = \begin{pmatrix} c_{1,1} & \cdots & c_{1,n} \\ \vdots & \ddots & \vdots \\ c_{n,1} & \cdots & c_{n,n} \end{pmatrix},$$

where $v_i = c_{1,i} \cdot w_1 + \cdots + c_{n,i} \cdot w_n$.

Indeed, the i-th column of $C_{\mathcal{W}\leftarrow\mathcal{V}}$ consists of the coordinates v_i in the basis $\mathcal{W}$, as expected. We're trying to go from basis $\mathcal{V}$ to basis $\mathcal{W}$, so we should definitely be representing old basis (namely, $v_1, \cdots, v_n$) in terms of $w_1, \cdots, w_n$.

Proposition 7.7. Suppose V is a finite-dimensional vector space over $\mathbb{F}$. Let $\mathcal{V} = (v_1, \cdots, v_n)$ and $\mathcal{W} = (w_1, \cdots, w_n)$ be two bases of V. Then, $C_{\mathcal{W}\leftarrow\mathcal{V}}$ is invertible, and $C^{-1}_{\mathcal{W}\leftarrow\mathcal{V}} = C_{\mathcal{V}\leftarrow\mathcal{W}}$.

The proof is obvious from the definition. Now that we have this machinery, we can finally prove the consistency of the definition:

Proof. Note that $\mathcal{M}_{\mathcal{W}}(T) = C_{\mathcal{W}\leftarrow\mathcal{V}} \cdot \mathcal{M}_{\mathcal{V}}(T) \cdot C_{\mathcal{V}\leftarrow\mathcal{W}}$, so

$$\begin{aligned} \det \mathcal{M}_{\mathcal{W}}(T) &= \det(C_{\mathcal{W}\leftarrow\mathcal{V}} \cdot \mathcal{M}_{\mathcal{V}}(T) \cdot C_{\mathcal{V}\leftarrow\mathcal{W}}) \\ &= \det C_{\mathcal{W}\leftarrow\mathcal{V}} \cdot \det \mathcal{M}_{\mathcal{V}}(T) \cdot \det C_{\mathcal{V}\leftarrow\mathcal{W}} \\ &= \det(C_{\mathcal{W}\leftarrow\mathcal{V}} \cdot C_{\mathcal{V}\leftarrow\mathcal{W}}) \cdot \det \mathcal{M}_{\mathcal{V}}(T) \\ &= \det I \cdot \det \mathcal{M}_{\mathcal{V}}(T) \\ &= \det \mathcal{M}_{\mathcal{V}}(T). \end{aligned}$$

Therefore, for any $T \in \mathcal{L}(V)$ and any two bases $\mathcal{V}$ and $\mathcal{W}$ of V, $\det \mathcal{M}_{\mathcal{V}}(T) = \det \mathcal{M}_{\mathcal{W}}(T)$. □

Proposition 7.8. Let $A \in \mathbb{F}^{n,n}$ be upper-triangular. Then,

$$\det A = A_{1,1} \cdots A_{n,n}.$$

Proof. In the definition, consider each term corresponding to some $\sigma \in S_n$. If $\sigma \neq (1, \cdots, n)$, then we must be going off the diagonal. But since we have to go through all n of them, they go in pairs: going above the diagonal once for some A_{i,σ_i} means you'll also need to go below the diagonal once, and vice versa. So, at least one of the A_{j,σ_j} is 0. Then, only the term with $\sigma = (1, \cdots, n)$ remains, and the proof is complete. □

7.2 The Characteristic Polynomial

The determinant allows us to arrive at the characteristic polynomial, a piece of machinery so potent it would be a waste not to introduce. The definition is simple for **complex** vector spaces:

Definition 7.9. Let V be a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. The characteristic polynomial of T, denoted as $\chi_T \in \mathcal{P}_{\dim V}(\mathbb{C})$, is

$$\chi_T(z) \coloneqq \det(z \cdot I - T),$$

which is a degree-$\dim V$, monic (i.e., leading coefficient equals 1) polynomial.

Proposition 7.10. Let V be a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Then, λ is an eigenvalue of T if and only if λ is a zero of χ_T.

Proof. Observe that $\chi_T(\lambda) = \det(\lambda \cdot I - T) = 0$ if and only if $\lambda \cdot I - T$ is not invertible by Corollary 7.4. By Corollary 4.20, this is equivalent to $\lambda \cdot I - T$, and thus $T - \lambda \cdot I$, being not injective. This is further equivalent to $\lambda \in E(T)$ by Proposition 6.3. $\square$

When we investigated polynomials earlier, we had Corollary 5.3 which provides the decomposition

$$\chi_T(z) = (z - \lambda_1)^{d_1} \cdots (z - \lambda_m)^{d_m}$$

for some $\lambda_1, \cdots, \lambda_m \in \mathbb{C}$ and $d_1, \cdots, d_m \in \mathbb{N}_+$ such that $d_1 + \cdots + d_m = n$ (note that $c = 1$ here). These $\lambda_1, \cdots, \lambda_m$ are precisely the eigenvalues of T, and $d_1, \cdots, d_m$ are said to be their (algebraic) multiplicities respectively.

The coefficients of χ_T also contains a lot of useful information, of which the most notable is the following:

Proposition 7.11. Let V be a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Suppose $\chi_T(z) = a_0 + \cdots + a_{n-1}z^{n-1} + z^n$. Then, $a_0 = (-1)^{\dim V} \cdot \det T$.

To prove this, we cleverly use the equivalence of the many forms of χ_T.

Proof. Note that $a_0 = \chi_T(0) = \det(0 \cdot I - T) = \det((-1) \cdot T) = (-1)^{\dim V} \cdot T$. $\square$

And if we look at the polynomial decomposition, we see that the constant terms comes from exactly one term in the decomposition: taking $-\lambda_i$ for each factor, where we still get that sign $(-1)^{\dim V}$. Therefore,

Corollary 7.12. Let V be a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Then, the determinant of T is the product of all eigenvalues of T, counted with their respective algebraic multiplicities.

7.3 The Trace

The definition of the trace of a matrix is straightforward: we just add up the diagonal entries.

Definition 7.13. Let $A \in \mathbb{F}^{n,n}$. The trace of A, denoted as $\operatorname{tr} A$, is defined as the sum of the diagonal entries of A; that is,

$$\operatorname{tr} A \coloneqq A_{1,1} + \cdots + A_{n,n}.$$

Even though there is not an immediate intuitive picture that we can attach to the trace, it still has many nice properties surprisingly. As is the case for determinants also, the trace turns out to underlie some very profound structure that is beyond the scope of these notes.

One important property is that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$—even though $AB \neq BA$ in general.

Proposition 7.14. Let $A, B \in \mathbb{F}^{n,n}$. Then, $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

We will again omit the proof: it's a useful exercise to show this from the definition, but too brute-force that it hides the amazingly coherent algebraic structure behind the trace, namely, the tensor product.

For this reason, we can similarly define the trace of an operator independent of the basis chosen.

Definition 7.15. Let V be a finite-dimensional vector space and $T \in \mathcal{L}(V)$. The trace of T, denoted as $\operatorname{tr} T$, is defined as $\operatorname{tr} \mathcal{M}(T)$ under any basis.

To show consistency, we simply apply the Proposition above.

Proof. Let $\mathcal{V}, \mathcal{W}$ be bases of V. Then,

$$\begin{aligned}
\operatorname{tr} \mathcal{M}_{\mathcal{W}}(T) &= \operatorname{tr}(C_{\mathcal{W} \leftarrow \mathcal{V}} \cdot \mathcal{M}_{\mathcal{V}}(T) \cdot C_{\mathcal{V} \leftarrow \mathcal{W}}) \\
&= \operatorname{tr}((C_{\mathcal{V} \leftarrow \mathcal{W}} \cdot C_{\mathcal{W} \leftarrow \mathcal{V}}) \cdot \mathcal{M}_{\mathcal{V}}(T)) \\
&= \operatorname{tr}(\mathcal{M}_{\mathcal{V}}(T)).
\end{aligned}$$

The proof is complete. $\square$

In the **complex** case, we can obtain an upper-triangular matrix of T under some basis, where the diagonal elements are precisely the eigenvalues (Corollary 6.15). Thus,

Proposition 7.16. Let V be a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Then, $\operatorname{tr} T$ equals the sum of all eigenvales of T, counted with their respective algebraic multiplicities.

Proof. Fix a basis under which $\mathcal{M}(T)$ is upper-triangular by Corollary 6.15; let $\lambda_1 = \mathcal{M}(T)_{1,1}, \cdots, \lambda_n = \mathcal{M}(T)_{n,n}$, where $n := \dim V$. Note that they are exactly all eigenvalue of T counted with their algebraic multiplicities, and $\operatorname{tr} T = \operatorname{tr} \mathcal{M}(T) = \lambda_1 + \cdots + \lambda_n$. $\square$

7.4 The Complexification of a Real Vector Space

So we have obtained many useful results about finite-dimensional complex vector spaces, but they won't hold at all for the real case (like the upper-triangularization). Is there a way to bridge this gap?

It turns out that we can nicely "complexify" any real vector space by observing how to get from $\mathbb{R}$ to $\mathbb{C}$: by taking $a + bi$ for $a, b \in \mathbb{R}$.

Definition 7.17. Let V be a real vector space. The complexification of V, denoted as $V_{\mathbb{C}}$, is the complex vector space associated with V defined as

$$V_{\mathbb{C}} := \{u + iv \mid u, v \in V\},$$

where $u + iv$ is a formal symbol defined by the pair $(u, v) \in V \times V$. As a vector space, the addition is defined component-wise from the product, and the scalar multiplication is defined as

$$(a + bi) \cdot (u + iv) := (a \cdot u - b \cdot v) + i(a \cdot v + b \cdot u)$$

for arbitrary $a + bi \in \mathbb{C}$ and $u + iv \in V_{\mathbb{C}}$.

The best part about the complexification is that any old basis of V is also a basis of $V_{\mathbb{C}}$! Even though we have twice as many "degrees of freedom," informally speaking, the complex scalars are sufficient to span the entire new space.

Proposition 7.18. Suppose V is a finite-dimensional real vector space and $v_1, \cdots, v_n$ is a basis of V. Then, $v_1, \cdots, v_n$ is a basis of $V_{\mathbb{C}}$.

P.S.: It's worth mentioning that any real vector $v \in V$ is to be interpreted as complex vector $v \in V_{\mathbb{C}}$ through "$v = v + i0_V$."

Proof. Clearly, $v_1, \cdots, v_n$ remains linearly independent, because $0_{\mathbb{R}} = 0_{\mathbb{C}}$. It remains to show that the list is spanning.

Suppose $u + iv \in V_{\mathbb{C}}$ is arbitrary with $u, v \in V$. Then, fix constants $a_1, \cdots, a_n$ and $b_1, \cdots, b_n$ such that $u = a_1 \cdot v_1 + \cdots + a_n \cdot v_n$ and $v = b_1 \cdot v_1 + \cdots + b_n \cdot v_n$. Then,

$$u + iv = (a_1 \cdot v_1 + \cdots + a_n \cdot v_n) + i(b_1 \cdot v_1 + \cdots + b_n \cdot v_n) = (a_1 + b_1 i) \cdot v_1 + \cdots + (a_n + b_n i) \cdot v_n.$$

Therefore, $v_1, \cdots, v_n$ is a linearly independent list in $V_{\mathbb{C}}$ that spans $V_{\mathbb{C}}$, so it is a basis of $V_{\mathbb{C}}$. $\square$

As a corollary, for a finite-dimensional real vector space, its complexification have the same dimension as the real vector space. But they're not isomorphic! The concept of vector space isomorphisms only makes sense for vector spaces over the same field.

Corollary 7.19. Let V be a finite-dimensional real vector space. Then,

$$\dim V_{\mathbb{C}} = \dim V.$$

So far, we still don't have what we need to transfer the complex stuff over to the real case. We need to also translate from operators on V to operators on $V_{\mathbb{C}}$.

Definition 7.20. Let V be a real vector space and $T \in \mathcal{L}(V)$. The complexification of T, denoted as $T_{\mathbb{C}} \in \mathcal{L}(V_{\mathbb{C}})$, is defined as

$$T_{\mathbb{C}}(u + iv) := T(u) + iT(v).$$

This definition is so natural that the matrix of a complexified operator equals the old matrix under the same basis, for an operator on a finite-dimensional real vector space.

Proposition 7.21. Let V be a finite-dimensional real vector space and $T \in \mathcal{L}(V)$. Under any basis,

$$\mathcal{M}(T) = \mathcal{M}(T_{\mathbb{C}}).$$

Proof. Fix a basis $v_1, \cdots, v_n$ of V and $V_{\mathbb{C}}$. Note that $T(v_i) = T_{\mathbb{C}}(v_i)$ by definition, so $\mathcal{M}(T) = \mathcal{M}(T_{\mathbb{C}})$. □

Because we defined the characteristic polynomial through matrices, the polynomials also agree:

Corollary 7.22. Let V be a finite-dimensional real vector space and $T \in \mathcal{L}(V)$. Then, $\chi_T = \chi_{T_{\mathbb{C}}}$.

Proof. Fix a basis $v_1, \cdots, v_n$ of V. By definition, for a real $z \in \mathbb{R}$, $\chi_T(z) = |z \cdot I_n - \mathcal{M}(T)| = |z \cdot I_n - \mathcal{M}(T_{\mathbb{C}}| = \chi_{T_{\mathbb{C}}}|_{\mathbb{R}}(z)$. That is, $\chi_{T_{\mathbb{C}}}|_{\mathbb{R}} - \chi_T$ is the constant zero polynomial, so the coefficients of $\chi_{T_{\mathbb{C}}}$ equal those of χ_T. □

As a result, the results for complex operators also hold for real operators!

Corollary 7.23. Let V be a finite-dimensional real vector space and $T \in \mathcal{L}(V)$. Then,

$$\begin{cases} \det T = \lambda_1 \cdots \lambda_n, \\ \operatorname{tr} T = \lambda_1 + \cdots + \lambda_n, \end{cases}$$

where $\lambda_1, \cdots, \lambda_n$ are the zeros of $\chi_{T_{\mathbb{C}}}$ counted with algebraic multiplicities.

8 Inner Product Spaces

The concept of a vector space is already so general and powerful, but we still can't do everything we normally can in $\mathbb{R}^n$. Specifically, because there is no concept of lengths or angles, we can't manipulate the geometry of a general vector space as we would in $\mathbb{R}^n$.

So what is the best tool to capture this concept? We could go with the length (a normed space), but a more general tool is the dot product in $\mathbb{R}^n$:

$$(x_1, \cdots, x_n) \cdot (y_1, \cdots, y_n) = x_1 y_1 + \cdots + x_n y_n.$$

Why is this such a big deal? Well, this is how we defined things like length (by taking the dot product of a vector with itself) and angles (do you remember that $x \cdot y = |x| \cdot |y| \cdot \cos\theta$ with θ being the angle between x and y?). And this concept can be readily formalized more generally.

Definition 8.1. An inner product space is a vector space V over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ equipped with an operator $\langle\cdot,\cdot\rangle \colon V \times V \to \mathbb{F}$ such that for any $u_1, u_2, u, v \in V$ and $c \in \mathbb{F}$,

- $\langle v, v\rangle \in [0, +\infty)$;
- $\langle v, v\rangle = 0 \Rightarrow v = 0$;
- $\langle c \cdot u_1 + u_2, v\rangle = c \cdot \langle u_1, v\rangle + \langle u_2, v\rangle$;

- $\langle u, v\rangle = \overline{\langle v, u\rangle}$,

where the bar denoted complex conjugation (which is the identity map on $\mathbb{R}$).

Obviously, the dot product in $\mathbb{R}^n$ makes the vector space $\mathbb{R}^n$ also an inner product space by this definition. But what the heck is the bar?

Well, we definitely need it if we want this concept to work in the complex case. The complex dot product is

$$(w_1, \cdots, w_n) \cdot (z_1, \cdots, z_n) = w_1\overline{z}_1 + \cdots + w_n\overline{z}_n.$$

This clearly fits the definition. But what if we used the old definition? It turns out the first property can't hold. Note that $\langle v, v\rangle$ won't necessarily be a real number (much less non-negative)! Can you come up with an example?

What's more, the second property also fails. For example, in $\mathbb{C}^2$, if we use the real vector space inner product, the vector $(1+i, 1-i)$ has

$$\text{“}\langle (1+i, 1-i), (1+i, 1-i)\rangle\text{”} = (1+i)^2 + (1-i)^2 = 1 + 2i - 1 + 1 - 2i - 1 = 0,$$

but $(1+i, 1-i) \neq 0$.

A sad consequence of this definition, though, is that the inner product is not linear in the second component for complex vector spaces (though this is the case in real vector spaces). While we still have $\langle u, v_1 + v_2\rangle = \langle u, v_1\rangle + \langle u, v_2\rangle$

I hope this can convince you that this definition is the way to go. Indeed, we get so many results from geometry, of which the most famous is the Pythagorean theorem. To establish this, we first need the notion of perpendicularity and length.

Definition 8.2. Let V be an inner product space and $u, v \in V$. Then, u is said to be perpendicular to v, denoted as $u \perp v$, if $\langle u, v\rangle = 0$.

While this is symmetric (that is, $u \perp v$ iff $v \perp u$), it is by no means an equivalence relation! Reflexivity clearly fails (for any nonzero vector), and so does transitivity. Consider $\mathbb{R}^3$ where $u = (1, 0, 0)$, $v = (1, 1, 0)$, and $z = (0, 0, 1)$. Even though $u \perp z$ and $v \perp z$, we don't have $u \perp v$.

Definition 8.3. Let V be an inner product space. The norm of V, denoted as $\|\cdot\| : V \to [0, +\infty)$, is defined as

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

for any $v \in V$.

By the way, I'll also slip in some properties of the norm: for any $c \in \mathbb{F}$ and $v, v_1, v_2 \in V$,

- (Positive definiteness) $\|v\| \geq 0$ and the equality holds if and only if $v = 0$;
- (Absolute homogeneity) $\|c \cdot v\| = |c| \cdot \|v\|$;
- (Triangle inequality) $\|v_1 + v_2\| \geq \|v_1\| + \|v_2\|$.

Note that the absolute value is not as trivial as you might think! This is pretty crucial especially for the complex case. The first two are obvious while proving the third is harder than it seems. We'll specifically prove the triangle inequality in just a moment.

For now, let's marvel at the Pythagorean theorem.

Theorem 8.4 (Pythagoras)**.** *Let V be an inner product space and $u, v \in V$. If $u \perp v$, then $\|u+v\|^2 = \|u\|^2 + \|v\|^2$.*

Proof. Observe that

$$\|u+v\|^2 = \langle u+v, u+v\rangle$$

$$
\begin{aligned}
&= \langle u, u\rangle + \cancel{\langle u, v\rangle} + \cancel{\langle v, u\rangle} + \langle v, v\rangle \\
&= \langle u, u\rangle + \langle v, v\rangle \\
&= \|u\|^2 + \|v\|^2 .
\end{aligned}
$$

The proof is complete. □

That was smooth. Now, we'll try to build towards the triangle inequality as promised, but we need some machinery.

In MATH 230, we talked about projections: for two vectors $u, v \in V$ with $v \neq 0$, we can write u as a sum of two vectors, one parallel to v and the other perpendicular to v. This also works for a general inner product space.

Theorem 8.5 (Orthogonal Decomposition)**.** *Suppose V is an inner product space and $u, v \in V$ with $v \neq 0$. Then, there exists $c \in \mathbb{F}$ and $w \in V$ with $w \perp v$ such that*

$$u = c \cdot v + w.$$

In particular, this holds for $c = \langle u, v\rangle / \langle v, v\rangle$ and $w = u - c \cdot v$.

The proof is straightforward, just verifying the desired properties hold.

Proof. Note that with $c = \langle u, v\rangle / \langle v, v\rangle$ and $w = u - c \cdot v$, we have

$$c \cdot v + w = c \cdot v + u - c \cdot v = u$$

and

$$\langle w, v\rangle = \langle u - c \cdot v, v\rangle = \langle u, v\rangle - c \cdot \langle v, v\rangle = \langle u, v\rangle - \frac{\langle u, v\rangle}{\langle v, v\rangle} \cdot \langle v, v\rangle = 0.$$

The proof is finished. □

The next big-name result is the Cauchy-Schwartz inequality. Those involved in math competition can truly appreciate how powerful this innocent formula looks.

Proposition 8.6 (Cauchy-Schwartz)**.** Let V be an inner product space and $u, v \in V$. Then,

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

where the equality holds if and only if u, v is linearly dependent.

Proof. We first prove the inequality. If $v = 0$, both sides equal 0, so the inequality holds trivially. Now suppose $v \neq 0$.

Let $\mathbb{F}$ be the field associated with V. By orthogonal decomposition (Theorem 8.5, fix $c \in \mathbb{F}$ and $w \in V$ with $w \perp v$ such that $u = c \cdot v + w$. Then,

$$
\begin{aligned}
\|u\|^2 &= \|c \cdot v\|^2 + \|w\|^2 && \text{(Theorem 8.4)} \\
&= |c|^2 \cdot \|v\|^2 + \|w\|^2 \\
&\geq \left| \frac{\langle u, v\rangle}{\langle v, v\rangle} \right|^2 \cdot \|v\|^2 && (\|w\|^2 \geq 0) \\
&= \frac{|\langle u, v\rangle|^2}{\langle v, v\rangle}.
\end{aligned}
$$

Therefore, $\|u\|^2 \cdot \|v\|^2 \geq |\langle u, v\rangle|^2$. Note that the equality holds iff $v = 0$ or, when $v \neq 0$, $w = 0$ (i.e., $u = c \cdot v$); that is, the equality holds if and only if u, v is linearly independent. □

It took us a while, but it was worth it. We're now ready to prove the triangle inequality.

Proposition 8.7 (Triangle Inequality)**.** Suppose V is a finite-dimensional vector space and $u, v \in V$. Then, $\|u+v\| \leq \|u\| + \|v\|$.

Proof. Observe that

$$\begin{aligned}
\|u+v\|^2 &= \langle u+v, u+v \rangle \\
&= \langle u, u \rangle + \langle u, v \rangle + \overline{\langle u, v \rangle} + \langle v, v \rangle \\
&= \langle u, u \rangle + 2 \operatorname{Re} \langle u, v \rangle + \langle v, v \rangle \\
&\leq \|u\|^2 + 2 \left|\langle u, v \rangle\right| + \|v\|^2 \\
&\leq \|u\|^2 + 2 \|u\| \cdot \|v\| + \|v\|^2 \\
&= (\|u\| + \|v\|)^2.
\end{aligned}$$

Therefore, $\|u\| + v \leq \|u\| + \|v\|$. □

A final note for the sake of completeness: the relationship between norms and inner products. As we have shown, every inner product $\langle \cdot, \cdot \rangle$ induces a norm $\|\cdot\| : x \mapsto \sqrt{\langle x, x \rangle}$. Can we get an inner product from a norm?

The answer is a qualified "yes." We first state two results that helps us evaluate an inner product from norms, without proof.

Proposition 8.8. Let V be a real inner product space. Then, for any $u, v \in V$,

$$\langle u, v \rangle = \frac{1}{4} \left(\|u+v\|^2 - \|u-v\|^2 \right).$$

Proposition 8.9. Let V be a complex inner product space. Then, for any $u, v \in V$,

$$\langle u, v \rangle = \frac{1}{4} \left(\|u+v\|^2 - \|u-v\|^2 + i \cdot \|u+iv\|^2 - i \cdot \|u-iv\|^2 \right).$$

This gives us a promising direction: for every normed space $(V, \|\cdot\|)$, we should be able to define an inner product in this way. The question is: is the resultant "inner product" really an inner product?

No! The norm may not satisfy the parallelogram equality: consider the ℓ_1 norm $\|\cdot\|_1$ on $\mathbb{R}^n$ defined by

$$\|(x_1, \cdots, x_n)\|_1 := |x_1| + \cdots + |x_n|$$

is indeed a norm, but the parallelogram equality does not hold (check this!). If we do try and define an "inner product" in this fashion, the parallelogram equality clearly fails for the inner product space, a contradiction. But otherwise, we do have an inner product induced by a norm!

8.1 Orthonomal Bases

Prototypical Example. $(1/\sqrt{2}, 1/\sqrt{2})$, $(-1/\sqrt{2}, 1/\sqrt{2})$ is an orthonormal basis of $\mathbb{R}^2$.

In a general finite-dimensional vector space V, any length-n linearly independent list of vectors is a basis, and there is no reasonable preference for one basis over another. But the geometry that the inner product introduces allows us to talk about perpendicularity, as we have seen, and it would be natural to ask for a basis like the standard basis on $\mathbb{R}^n$: a pairwise perpendicular one where each basis vector has length 1.

Why is this interesting? Well, a lot of results ensue from this perpendicularity, which adds so much richer structure to this space, and allows us to start quantifying the geometry of this space. A particularly appealing aspect of such a basis is that we can directly calculate the coordinate of a vector, which isn't possible in a vector space! Consider a vector $v = (x_1, \cdots, x_n) \in \mathbb{R}^n$. In order to get the first coordinate x_1, we can simply do

$$x_i = \langle v, e_1 \rangle.$$

Before I get too excited, let's start with the definition.

Definition 8.10. Suppose V is an inner product space. A list of vectors $e_1, \cdots, e_k \in V$ is said to be orthonormal if

$$\langle e_i, e_j \rangle = \delta_{i,j}.$$

If V is finite-dimensional and $e_1, \cdots, e_n$ is a basis, then $e_1, \cdots, e_n$ is also called an orthonormal basis.

Just as a reminder, the $\delta_{i,j}$ notation is called the Kronecker delta. It is 1 when $i = j$ and 0 otherwise, straightforward enough. This is just a concise notation that captures both the length-1 property and the pairwise orthogonal property.

Now, let's establish some results we mentioned so far, like coordinates and stuff. We'll also refrain from adding the finite-dimensional condition, whenever possible, to make the discussion more general. After all, many useful spaces are infinite-dimensional (an important example is $C^k(\mathbb{R})$ for $k \in \mathbb{Z}_{\geq 0}$, the space of k-times continuously differentiable functions from $\mathbb{R}$).

Theorem 8.11. *Suppose V is an inner product space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and $e_1, \cdots, e_k \in V$ is an orthonormal list. Then,*

- $\|c_1 \cdot e_1 + \cdots + c_k \cdot e_k\|^2 = |c_1|^2 + \cdots + |c_k|^2$ *for all* $c_1, \cdots, c_k \in \mathbb{F}$*;*
- $e_1, \cdots, e_k$ *is linearly independent;*
- $v = \langle v, e_1 \rangle \cdot e_1 + \cdots + \langle v, e_k \rangle \cdot e_k$ *for all* $v \in \operatorname{span}(e_1, \cdots, e_k)$.

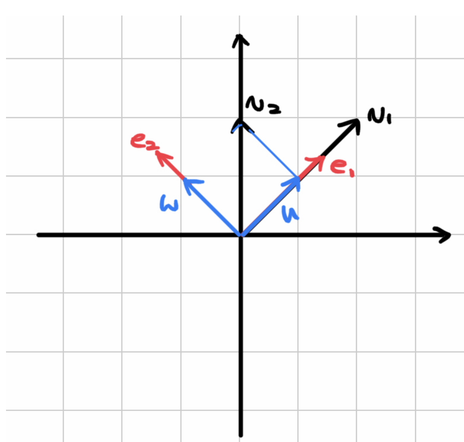
Proof. The first item follows from repeated applications of the Pythagorean theorem. Note that if $v \in V$ is perpendicular to any of $e_1, \cdots, e_k$, then v is perpendicular to any linear combination $(c_1 \cdot e_1 + \cdots + c_k \cdot e_k)$ for all $c_1, \cdots, c_k \in \mathbb{F}$. Thus,

$$\begin{aligned} \|c_1 \cdot e_1 + \cdots + c_k \cdot e_k\|^2 &= \|c_1 \cdot e_1\|^2 + \|c_2 \cdot e_2 + \cdots + c_k \cdot e_k\|^2 \\ &= |c_1|^2 + \|c_2 \cdot e_2 + \cdots + c_k \cdot e_k\|^2 \\ &= \vdots \\ &= |c_1|^2 + \cdots + |c_k|^2 . \end{aligned}$$

For item 2, suppose $c_1, \cdots, c_k \in \mathbb{F}$ are such that $c_1 \cdot e_1 + \cdots + c_k \cdot e_k = 0$. Then, by item 1, $|c_1|^2 + \cdots + |c_k|^2 = 0$, which implies $c_1 = \cdots = c_k = 0$; that is, $e_1, \cdots, e_k$ is linearly independent.

Lastly, suppose $v = c_1 \cdot e_1 + \cdots + e_k \cdot e_k$. Then, for any $i \in \{1, \cdots, k\}$, $\langle v, e_i \rangle = c_1 \cdot \langle e_1, e_i \rangle + \cdots + c_k \cdot \langle e_k, e_i \rangle = c_i$ indeed. □

Let's now take a look at the famed Gram-Schmidt process, an algorithm that "orthonormalizes" all linearly independent lists. Say we have $v_1 = (1, 1)$ and $v_2 = (0, 1)$ in $\mathbb{R}^2$. Clearly, they can span the entire plane. The process will first divide v_1 by



its length to get us a unit vector:

$$e_1 := v_1/\|v_1\| = (1/\sqrt{2}, 1/\sqrt{2}).$$

Now, v_2 isn't perpendicular to e_1, but we can decompose it as $v_2 = u + w$, where u is parallel to e_1 and w perpendicular to e_1 (Theorem 8.5). Then, taking just the w part works, so we do

$$e_2 := w/\|w\| = (-1/\sqrt{2}, 1/\sqrt{2}).$$

We'll present this result in terms of a theorem that justifies some of its useful theoretical properties.

Theorem 8.12 (Gram-Schmidt)**.** *Let V be an inner product space and suppose $v_1, \cdots, v_n \in V$ form a linearly independent list. Let*

$$\begin{aligned}
u_1 &= v_1, & e_1 &= u_1/\|u_1\|, \\
u_2 &= v_2 - \langle v_2, e_1\rangle \cdot e_1, & e_2 &= u_2/\|u_2\|, \\
u_3 &= v_3 - \langle v_3, e_1\rangle \cdot e_1 - \langle v_3, e_2\rangle \cdot e_2, & e_3 &= u_3/\|u_3\|, \\
&\vdots & &\vdots \\
u_n &= v_n - \langle v_n, e_1\rangle \cdot e_1 - \cdots - \langle v_3, e_n\rangle \cdot e_n, & e_n &= u_n/\|u_n\|.
\end{aligned}$$

Then, $e_1, \cdots, e_n$ is an orthonormal list and $\mathrm{span}(e_1, \cdots, e_i) = \mathrm{span}(v_1, \cdots, v_i)$ *for all $i \in \{1, \cdots, n\}$.*

Proof. We perform induction on i.

Base case. Note that $e_1 = u_1/\|u_1\|$ has norm $\|e_1\| = \|u_1/\|u_1\|\| = |\|u_1\|| / \|u_1\| = 1$. This is valid because $v_1 \neq 0$ (because of linear independence). Then, e_1 is an orthonormal list and $\mathrm{span}(e_1) = \mathrm{span}(v_1)$.

Inductive case. Suppose now that $e_1, \cdots, e_{i-1}$ is an orthonormal list with $\mathrm{span}(e_1, \cdots, e_{i-1}) = \mathrm{span}(v_1, \cdots, v_{i-1})$ for some $i \in \{2, \cdots, n\}$. Similarly, linear independence implies $v_i \notin \mathrm{span}(v_1, \cdots, v_{i-1}) = \mathrm{span}(e_1, \cdots, e_{i-1})$, so $u_i \neq 0$ by Theorem 8.11. Then e_i also has length 1.

It remains to verify perpendicularity. Indeed, for any $j \in \{1, \cdots, i-1\}$,

$$\begin{aligned}
\langle e_i, e_j\rangle &= \frac{1}{\|u_i\|} \cdot \langle v_i - \langle v_i, e_1\rangle \cdot e_1 - \cdots - \langle v_i, e_{i-1}\rangle \cdot e_{i-1}, e_j\rangle \\
&= \frac{1}{\|u_i\|} \cdot \left(\langle v_i, e_j\rangle - \cancel{\langle v_i, e_1\rangle \cdot \langle e_1, e_j\rangle} - \cdots - \langle v_i, e_j\rangle \cdot \langle e_j, e_j\rangle - \cdots - \cancel{\langle v_i, e_{i-1}\rangle \cdot \langle e_{i-1}, e_j\rangle}\right) \\
&= 0.
\end{aligned}$$

The proof is complete. □

Not only is the Gram-Schmidt process useful in a wide variety of applications, this result also has very useful corollaries theoretically.

Corollary 8.13. Let V be a finite-dimensional inner product space. Then,

- V has an orthonormal basis;
- Any orthonormal list of vectors in V can be extended to an orthonormal basis of V;

The proof are trivial, but the implication is not: the finite-dimensionality of an inner product space now implies the existence of not only an arbitrary basis, but an orthonormal one. Another even more useful result is presented as follows; it is so special that it deserves its own name:

Theorem 8.14 (Schur)**.** *Let V be a non-zero, finite-dimensional complex inner product space. Then, any linear operator $T \in \mathcal{L}(V)$ has an upper-triangular matrix under some orthonormal basis of V.*

Proof. Because V is a non-zero, finite-dimensional complex vector space, $\mathcal{M}_{\mathcal{V}}(T)$ is upper-triangular under some basis $\mathcal{V} := v_1, \cdots, v_n$ of V by Theorem 6.13. Then, Proposition 6.12 implies that $\text{span}(v_1, \cdots, v_i)$ is T-invariant for all $i \in \{1, \cdots, n\}$. Then, after applying the Gram-Schmidt process to $v_1, \cdots, v_n$, we have an orthonormal basis $\mathcal{E} := e_1, \cdots, e_n$ of V. Now, for all $i \in \{1, \cdots, n\}$, $\text{span}(e_1, \cdots, e_i) = \text{span}(v_1, \cdots, v_i)$ is T-invariant, so $\mathcal{M}_{\mathcal{E}}(T)$ is upper-triangular. □

Now let's turn to the topic of linear functionals, or the dual space. We are concerned with this question in the setting of a finite-dimensional inner product space. We've talked about how a linear functional $\phi \in V^* = \mathcal{L}(V, \mathbb{F})$ acts like a row vector for $V = \mathbb{F}^n$, in the sense that it is linear and has codomain $\mathbb{F}$. But there's a crucial missing piece: can we actually write any linear functional ϕ as a row vector, which is the transpose of some column vector $u \in V$? That is, we want to see if there's a $u \in V$ so that $\phi(v) =$ "$u^\top v$" $= \langle v, u\rangle$ for all v. The Riesz representation theorem assures us that this is the case. What's more, it also gives a way to explicitly calculate this u.

Theorem 8.15 (Riesz representation)**.** *Let V be a finite-dimensional inner product space. Then, for any linear functional $\phi \in V^*$, there exists a unique vector $u \in V$ such that*

$$\phi(v) = \langle u, v\rangle \qquad \textit{for all } v \in V.$$

Further, for any given orthonormal basis $e_1, \cdots, e_n$ of V,

$$u = \overline{\phi(e_1)} \cdot e_1 + \cdots + \overline{\phi(e_n)} \cdot e_n.$$

One thing to note is that even though u is written in different ways under different orthonormal bases, it's still the same vector! Now, although we can just plug back u to check this holds, we'll still spell out how we got there.

Proof. To show existence, we suppose $u \in V$ is such that $\phi(v) = \langle u, v\rangle$ for all $v \in V$ and argue that an instance of such u can be found explicitly. Let $e_1, \cdots, e_n$ be an orthonormal basis of V. Then, for all $i \in \{1, \cdots, n\}$,

$$\phi(e_i) = \langle e_i, u\rangle = \overline{\langle u, e_i\rangle};$$

that is, the i-th coordinate of u is $\langle u, e_i\rangle = \overline{\phi(e_i)}$. We now show that $u = \overline{\phi(e_1)} \cdot e_1 + \cdots + \overline{\phi(e_n)} \cdot e_n$ works: for any $v = c_1 \cdot e_1 + \cdots + c_n \cdot e_n$,

$$\langle v, u\rangle = c_1 \cdot \langle e_1, u\rangle + \cdots + c_n \cdot \langle e_n, u\rangle = c_1 \cdot \phi(e_1) + \cdots + c_n \cdot \phi(e_n) = \phi(v).$$

We proceed to show uniqueness. Suppose $u_1, u_2 \in V$ are such that $\phi(v) = \langle v, u_1\rangle = \langle v, u_2\rangle$ for all $v \in V$. Then, $\langle v, u_1 - u_2\rangle$=0, so $\langle u_1 - u_2, e_i\rangle = 0$ for all $i \in \{1, \cdots, n\}$. So, $u_1 - u_2 = 0 \cdot e_1 + \cdots + 0 \cdot e_n = 0$, and thus $u_1 = u_2$. □

8.2 Orthogonal Complements

Prototypical Example. In $\mathbb{R}^3$, the orthogonal complement of the y-axis is the xz-plane.

In geometry, the definition of orthogonality extends from lines (vectors) to geometric objects (planes). In 3d, for example, we talk about a line perpendicular to a plane. If two points are symmetric with respect to a line, then the points are on the same perpendicular plane. We generalize this idea to abstract inner product spaces.

Definition 8.16. Suppose V is an inner product space and let $U \subseteq V$. The orthogonal complement of U, denoted as $U^\perp \subseteq V$, is defined as the collection of vectors that are perpendicular to all vectors in U; that is,

$$U^\perp := \{v \in V \mid \forall u \in V, v \perp u\}.$$

We give some immediate results of this definition.

Proposition 8.17. Suppose V is an inner product space and let $U \subseteq V$. Then,

- $U^\perp$ is a subspace of V;
- $\{0\}^\perp = V$;
- $V^\perp = \{0\}$;
- $U \cap U^\perp = \{0\}$.

Proof. Let $\mathbb{F}$ be the field associated with V. The first item is obvious: if $v_1, v_2 \in U^\perp$, then for any $c \in \mathbb{F}$, $\langle c \cdot v_1 + v_2, u\rangle = c \cdot \langle v_1, u\rangle + \langle v_2, u\rangle = 0$.

We move on to the second and third statements. By direct calculation, $\{0\}^\perp = \{v \in V \mid \langle v, 0\rangle = 0\} = V$ and $V^\perp = \{v \in V \mid \forall u \in V, v \perp u\} = \{0\}$.

Lastly, note that if $u \in U \cap U^\perp$, then $u \perp u$; that is, $\langle u, u\rangle = 0$, so $u = 0$. Indeed, 0 is in both subspaces. □

While U can be just a subset, we almost always talk about U as a subspace. But it doesn't hurt to have the above definitions in slightly greater generality.

The third item tells us that the sum of U and $U^\perp$ is a direct sum. We might wonder: what is this sum? Well, in the prototypical example, we know that the y-axis and the xz-plane direct-sums to the entire space $\mathbb{R}^3$. Is this always true?

Theorem 8.18. *Suppose V be an inner product space and let U be a finite-dimensional subspace of V. Then, $U \oplus U^\perp = V$.*

Proof. Fix an orthonormal basis $e_1, \cdots, e_n$ of U. For any $v \in V$, define $u = \langle v, e_1\rangle \cdot e_1 + \cdots + \langle v, e_n\rangle \cdot e_n \in U$. It suffices to show that $v - u \in U^\perp$. Similar to our analysis of the Gram-Schmidt process, for each basis vector e_i with $i \in \{1, \cdots, n\}$, we have $\langle v - u, e_i\rangle = \langle v - \langle v, e_i\rangle \cdot e_i, e_i\rangle = 0$. Then, $(v - u)$ is perpendicular to all basis vectors of U and hence the entirety of U, and thus $v - u \in U^\perp$. □

Corollary 8.19. Suppose V is a finite-dimensional inner product space and let U be a subspace of V. Then,

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

Another observation in $\mathbb{R}^3$ generalizes (kind of) well to general inner product spaces: the orthogonal complement of the y-axis is the xz-plane, and the orthogonal complement of the xz-plane is back to the y-axis. Is this true in general? Do we have $(U^\perp)^\perp = U$?

This is indeed the case for a finite-dimensional subspace U.

Proposition 8.20. Suppose V is an inner product space and let U be a finite-dimensional subspace of V. Then, $(U^\perp)^\perp = U$.

The approach to this proof is the following: to show the set equality, we show the subset relation in both directions. Unravel what the definition mean like peeling an onion layer-by-layer, and we'll get to what we need.

Proof. $U \subseteq (U^\perp)^\perp$. This direction is straightforward. For any $u \in U$ and $v \in U^\perp$, $u \perp v$ by definition; that is, u is perpendicular to the entirety of $U^\perp$, and hence $u \in (U^\perp)^\perp$.

$(U^\perp)^\perp \subseteq U$. We need to show that if v is perpendicular to the entirety of $U^\perp$, then $v \in U$. Suppose $v \perp v'$ for any $v' \in U^\perp$. By the direct sum decomposition, we write $v = u + w$, where $u \in U$ and $w \in U^\perp$. Now, v can also be written uniquely as a sum of vectors from $U^\perp$ and $(U^\perp)^\perp$ respectively. The uniqueness implies the first vector must be w and the second $u \in U \subseteq (U^\perp)^\perp$. Then, $w = v - u$, where both $v, u \in (U^\perp)^\perp$, so $w \in (U^\perp)^\perp$ by linear closure. Because $w \in U^\perp \cap (U^\perp)^\perp$, $w = 0$. Thus, $v = u \in U$ as required. □

This isn't true for an infinite-dimensional subspace U! These failures are typically found in analysis, but I'll give a quick example. Consider the space $l^2 \subset \mathbb{R}^\infty$ of square-summable sequences:

$$l^2 := \left\{ (x_1, x_2, \cdots) \in \mathbb{R}^\infty \quad : \quad \sum_{i=1}^{\infty} x_i^2 < \infty \right\}.$$

The convergence allows us to define an inner product

$$\langle (x_1, x_2, \cdots), (y_1, y_2, \cdots) \rangle := \sum_{i=1}^{\infty} x_i y_i.$$

Let's consider the subspace U of sequences with only finitely many non-zero elements. Note that $e_1 := (1, 0, 0, \cdots)$, $e_2 := (0, 1, 0, \cdots)$, $e_3 := (0, 0, 1, \cdots,), \cdots$ is an arbitrarily long independent list in U, so U is infinite-dimensional. Now, $U^\perp = \{0\}$ because any $v \in U^\perp$ must be perpendicular to all of U, and in particular the "one-hot" sequences:

$$v = (v_1, v_2, \cdots) \in U^\perp \Rightarrow \forall i \in \mathbb{N}_+, \langle v, e_i \rangle = v_i = 0 \Rightarrow v = 0.$$

Then, $\{0\}^\perp = V$; that is, $(U^\perp)^\perp = V \neq U$.

8.3 Orthogonal Projections

Prototypical Example. Drawing a cube on a piece of paper. This projects each point on the cube to a plane.

We've frequently used the concept of

$$u := \sum_{i=1}^{n} \langle v, e_i \rangle \cdot e_i$$

for an orthonormal basis $e_1, \cdots, e_n$ of some subspace U. Note that when $\dim U = 1$ (or $n = 1$), we recover the orthogonal decomposition, where $v = u + w$ for some $u \in U$ and $w \in U^\perp$. Let's generalize this further.

Because we have $U \oplus U^\perp = V$ for $\dim U < \infty$, we can do this.

Definition 8.21. Suppose V is an inner product space and U is a finite-dimensional subspace. For each $v \in V$, we have a unique decomposition

$$v = u + w$$

for $u \in U$ and $w \in U^\perp$. Define the orthogonal projection of v onto U, denoted as $P_U(v)$, as this unique u. This defines a map $P_U : V \to V$.

This definition now allows us to project onto not just a line, but a plane, hyperplane, etc. In many calculations, we still want to use the basis. We'll now formalize this notion.

Proposition 8.22. Suppose V is an inner product space and U is a finite-dimensional subspace. If $e_1, \cdots, e_n$ is an orthonormal basis of U, then

$$P_U(v) = \langle v, e_1 \rangle \cdot e_1 + \cdots + \langle v, e_n \rangle \cdot e_n$$

for all $v \in V$.

For this proof, note that we already have a u. Because of the uniqueness guaranteed by the direct sum, it only remains to show that the $w = v - P_U(v)$ is in $U^\perp$.

Proof. Let $u = \langle v, e_1 \rangle \cdot e_1 + \cdots + \langle v, e_n \rangle \cdot e_n$ and $w = v - u$. It suffices to show $w \in U^\perp$; that is, $w \perp e_i$ for all $i \in \{1, \cdots, n\}$. Observe that

$$\langle w, e_i \rangle = \langle v, e_i \rangle - \sum_{j=1}^{n} \left\langle v, e_j \right\rangle \cdot \left\langle e_j, e_i \right\rangle = \langle v, e_i \rangle - \sum_{j=1}^{n} \left\langle v, e_j \right\rangle \cdot \delta_{i,j} = \langle v, e_i \rangle - \langle v, e_i \rangle = 0.$$

Then, the linearity of the inner product implies that w is perpendicular to all of U, and hence $w \in U^\perp$. □

Here are some immediate properties of the orthogonal projection:

Proposition 8.23. Suppose V is an inner product space and U is a finite-dimensional subspace. Then,

- $P_U \in \mathcal{L}(V)$ (that is, P_U is linear);
- $\operatorname{null} P_U = U^\perp$;
- $\operatorname{range} P_U = U$;
- $P_U|_U = I$;
- $P_U^2 = P_U$;
- $P_U + P_{U^\perp} = I$;
- $\|P_U(v)\| \leq \|v\|$ for all $v \in V$.

Proof. Fix an orthonormal basis $e_1, \cdots, e_n$ of U.

The first item is obvious from the linearity of the expression from Proposition 8.22.

To show the second item, we show inclusion in both directions. For $\subseteq$, let $P_U(v) = 0$ for some $v \in V$. Then the coordinates $\langle v, e_i \rangle = 0$ for all $i \in \{1, \cdots, n\}$. Thus, v is perpendicular to the entirety of U, and hence $v \in U^\perp$. For $\supseteq$, let $v \in U^\perp$. In particular, $v \perp e_i$ for any $i \in \{1, \cdots, n\}$. Then,

$$P_U(v) = \langle v, e_1 \rangle \cdot e_1 + \cdots + \langle v, e_n \rangle \cdot e_n = 0 \cdot e_1 + \cdots + 0 \cdot e_n = 0.$$

The third through fifth items follow also from the expression above.

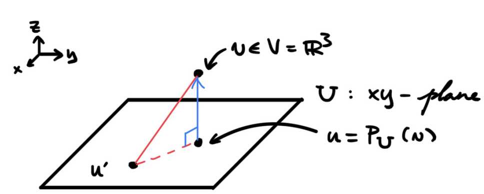
The sixth item can be shown by observing the direct sum. An arbitrary $v \in V$ admits a unique representation $v = u + w$ for $u \in U$ and $w \in U^\perp$. This is also a unique sum of elements $w \in U^\perp$ and $u \in U = (U^\perp)^\perp$. Then, by definition, $w = P_{U^\perp}(v)$. Then, $P_U(v) + P_{U^\perp}(v) = v$ for all v, and hence $P_U + P_{U^\perp} = I$.

The last item is a corollary of the Pythagoren theorem (Theorem 8.4). For an arbitrary $v \in V$, $v = u + w$ where $u = P_U(v) \in U$ and $w \in U^\perp$. Then, $u \perp w$ in particular, and hence

$$\|v\|^2 = \|u\|^2 + \|w\|^2 \geq \|u\|^2,$$

so $\|u\| = \|P_U(v)\| \leq \|v\|$. □

A big property of the projection is that it gives a **minimizer**: the projection u of v is the closest point in U to v.



In the example above, we see that $u = P_U(v)$ is the closest point in U to v. Any other u' would be able to form a right triangle, and where $\|u' - v\|$ is the hypotenuse that's longer than the "leg" of length $\|u - v\|$. This is basically the proof!

Theorem 8.24. *Suppose V is an inner product space and U is a finite-dimensional subspace. Then, for all $v \in V$,*

$$P_U(v) = \operatorname*{arg\,min}_{u \in U} \|u - v\|.$$

Proof. For any $v \in V$ and $u \in U$,

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

Thus, $\|P_U(v) - v\|^2 \leq \|u - v\|^2$ for any $u \in V$. The proof is complete. □

Perhaps the most useful application of this theoretical tool is in Fourier series. Let $V = C^0([-\pi, \pi])$ be the set of continuous function from $[-\pi, \pi]$ to $\mathbb{R}$, which is an inner product under

$$\langle f, g \rangle := \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) g(x) \, \mathrm{d}x.$$

Let's consider some basis functions, expressed in terms of the input x:

$$\frac{1}{\sqrt{2}}, \quad \sin x, \quad \cos x, \quad \sin 2x, \quad \cos 2x, \quad \cdots .$$

It's easy to verify that they're pairwise orthogonal with some (nasty) integral calculations. In fact, they are orthonormal. This allows us to define subspaces

$$U_n = \operatorname{span}(x \mapsto 1/\sqrt{2}, \quad x \mapsto \sin x, \quad x \mapsto \cos x, \quad \cdots, \quad x \mapsto \sin(nx), \quad x \mapsto \cos(nx))$$

of dimension $(2n + 1)$. The (finite) Fourier series is then the orthogonal projection f_0 of some given function $f \in V$:

$$f_0(x) = a_0/\sqrt{2} + \sum_{i=1}^{n} a_n \cos(nx) + \sum_{i=1}^{n} b_n \sin(nx),$$

where

$$\begin{cases} a_0 = \left\langle f, x \mapsto 1/\sqrt{2} \right\rangle = \dfrac{1}{\sqrt{2}\pi} \displaystyle\int_{-\pi}^{\pi} f(x) \, \mathrm{d}x, \\ a_n = \langle f, x \mapsto \cos(nx) \rangle = \dfrac{1}{\pi} \displaystyle\int_{-\pi}^{\pi} f(x) \cos(nx) \, \mathrm{d}x, \\ b_n = \langle f, x \mapsto \sin(nx) \rangle = \dfrac{1}{\pi} \displaystyle\int_{-\pi}^{\pi} f(x) \sin(nx) \, \mathrm{d}x. \end{cases}$$

Now, the orthogonal projection says that these coefficient are the best for f_0 in terms of making

$$\int_{-\pi}^{\pi} (f(x) - f_0(x))^2 \, \mathrm{d}x,$$

the (mean) squared error of f and f_0, the smallest.

8.4 Adjoints and Self-Adjoint Operators

Prototypical Example. The adjoint of $\begin{pmatrix} 1-i & 1 \\ 2 & 1+i \end{pmatrix} \in \mathcal{L}(\mathbb{C}^2)$ is $\begin{pmatrix} 1+i & 2 \\ 1 & 1-i \end{pmatrix} \in \mathcal{L}(\mathbb{C}^2)$.

Earlier, we have seen a first attempt at formalizing an equivalent of a (conjugate) transpose of a matrix. Let $T \in \mathcal{L}(V, W)$, for which there is a natural $T' \in \mathcal{L}(W', V')$ by defining

$$\forall \psi \in W', \qquad T'(\psi) := \psi \circ T.$$

For any given linear functional $\psi \in W'$, it is a function that takes a vector from W to a scalar. Now, chaining it after T, we get a functional in V' as the output of T': with input $v \in V$, we take $\psi(T(v)) \in \mathbb{F}$ as the output of this functional.

While convoluted, this is a really nice definition algebraically. But it would be better if we have something from W to V like an actual conjugate transpose, instead of mapping W' to V'. Luckily, in finite-dimensional cases, the Riesz representation theorem (Theorem 8.15) allows us to fix a one-to-one correspondence between vectors in V (resp. W) and functionals in V' (resp. W').

We could chain the Riesz representation with the definition of the dual map, and get an even worse-looking definition. But some careful calculations reveal that we can define them alternatively, in a more succinct and elegant form:

Definition 8.25. Suppose V and W are finite-dimensional inner product spaces. To each linear map $T \in \mathcal{L}(V, W)$ is associated uniquely its adjoint $T^*: W \to V$, such that

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

for all $v \in V$ and $w \in W$.

We'll first establish that this definition is valid. Well, fixing a particular $w \in W$, $\langle T(\cdot), w\rangle$ is a linear functional in V', so the Riesz representation theorem implies that there is some $v^* \in V$ so that $\langle \cdot, v^*\rangle$ is this particular functional. The arbitrary choice for v establishes the equality of the functionals. This v^* is called $T^*(w)$, which exists for any $w \in W$ and is unique.

Even though a transpose is a matrix and hence a linear map, we still need to check this explicitly for the adjoint.

Proposition 8.26. Suppose V and W are finite-dimensional inner product spaces over $\mathbb{F}$ and $T \in \mathcal{L}(V, W)$. Then, $T^* \in \mathcal{L}(W, V)$.

Proof. Let $c \in \mathbb{F}$, $v \in V$, and $w_1, w_2 \in W$. Then,

$$\langle v, T^*(c \cdot w_1 + w_2)\rangle = \langle T(v), c \cdot w_1 + w_2\rangle = \overline{c}\cdot\langle T(v), w_1\rangle + \langle T(v), w_2\rangle = \overline{c}\cdot\langle v, T^*(w_1)\rangle + \langle v, T^*(w_2)\rangle = \langle v, c \cdot T^*(w_1) + T^*(w_2)\rangle .$$

Because v is arbitrary, we conclude that $T^*(c \cdot w_1 + w_2) = c \cdot T^*(w_1) + T^*(w_2)$. □

We'll first state some general properties of the adjoint, which are also the property of the conjugate adjoint of a complex matrix.

Proposition 8.27. Suppose U, V, W are finite-dimensional inner product spaces over $\mathbb{F}$. Then,

- $(c \cdot S + T)^* = \overline{c} \cdot S^* + T^*$ for all $c \in \mathbb{F}$ and $S, T \in \mathcal{L}(V, W)$;
- $(T^*)^* = T$ for all $T \in \mathcal{L}(V, W)$;
- $(ST)^* = T^*S^*$ for all $S \in \mathcal{L}(V, W)$ and $T \in \mathcal{L}(U, V)$.

Proof. We use the fact that $\langle v, v_1\rangle = \langle v, v_2\rangle$ for all $v \in V$ implies $v_1 = v_2$, because subtracting the two sides implies $(v_1 - v_2)$ is orthogonal to any v and hence $(v_1 - v_2) \in V^\perp = \{0\}$.

For the first item, note that for all $v \in V$ and $w \in W$,

$$\langle v, (c \cdot S + T)^*(w)\rangle = \langle c \cdot S(v) + T(v), w\rangle = c \cdot \langle S(v), w\rangle + \langle T(v), w\rangle = c \cdot \langle v, S^*(w)\rangle + \langle v, T^*(w)\rangle = \langle v, (\overline{c} \cdot S^* + T^*)(w)\rangle .$$

To show the second, observe that for all $v \in V$ and $w \in W$,

$$\langle v, (T^*)^*(w)\rangle = \langle T^*(v), w\rangle = \langle v, T(w)\rangle .$$

Lastly, it is straightforward that for all $u \in U$ and $w \in W$,

$$\langle u, (ST)^*(w)\rangle = \langle S(T(u)), w\rangle = \langle T(u), S^*(w)\rangle = \langle u, T^*(S^*(w))\rangle .$$

The proof is complete. □

The adjoint captures the null space and the range well, and we have the following theoretically useful results.

Proposition 8.28. Suppose V and W are finite-dimensional inner product spaces and let $T \in \mathcal{L}(V, W)$. Then,

- $\operatorname{null} T^* = (\operatorname{range} T)^\perp$;
- $\operatorname{range} T^* = (\operatorname{null} T)^\perp$;
- $\operatorname{null} T = (\operatorname{range} T^*)^\perp$;
- $\operatorname{range} T = (\operatorname{null} T^*)^\perp$.

Proof. To show the first item, note that for any $w \in W$,

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

Taking the orthogonal complement of both sides gives item 2, noting that $(U^\perp)^\perp = U$.

Substituting $T \leftarrow T^*$ and noting $(T^*)^* = T$ gives the last two items. □

Now that we have some understanding about the basic properties of the adjoint, let's now show that this corresponds to the conjugate transpose of a matrix.

But before we dive in, there's a catch: we have to have orthonormal bases. This innocent assumption is by no means trivial: under non-orthonormal bases, $\mathcal{M}(T^*) \neq \overline{\mathcal{M}(T)}^\top$ in general! Let's first state the result, and then take a look at why orthonormal bases are such a big deal.

Proposition 8.29. Let V and W be finite-dimensional inner product spaces over $\mathbb{F}$ and suppose $T \in \mathcal{L}(V, W)$. Then, for any orthonormal bases $\mathcal{V}$ of V and $\mathcal{W}$ of W,

$$\mathcal{M}_{\mathcal{W},\mathcal{V}}(T^*) = \overline{\mathcal{M}_{\mathcal{V},\mathcal{W}}(T)}^\top.$$

Proof. Let $\mathcal{V} = e_1, \cdots, e_n$ and $\mathcal{W} = f_1, \cdots, f_m$ be orthonormal bases of V and W respectively. Then, for each $i \in \{1, \cdots, m\}$, the i-th column of $\mathcal{M}(T^*)$ holds the coefficients of $T^*(f_i) \in V$ under the basis $e_1, \cdots, e_n$. For any $j \in \{1, \cdots, n\}$,

$$\mathcal{M}(T^*)_{j,i} = \left\langle T^*(f_i), e_j \right\rangle = \left\langle f_i, T(e_j) \right\rangle = \overline{\left\langle T(e_j), f_i \right\rangle} = \overline{\mathcal{M}(T)_{i,j}}.$$

The proof is complete. □

Note how we had to use the orthonormal basis to get the coefficients: this is can't be done simply with the inner product. Under only an orthonormal basis do we have $\mathcal{M}(v)^\top \mathcal{M}(w) = \langle v, w \rangle$.

8.5 Self-Adjoint Operators

Prototypical Example. Conjugate symmetric matrices in $\mathbb{C}^n$, or the operator of any observable in elementary quantum mechanics.

In an inner product space, we can finally characterize what conjugate symmetric matrices really are: they are self-adjoint operators. By the way, if $T \in \mathcal{L}(V, W)$ with $V = W$, then the operators T and T^* don't even have the same domain and hence cannot possibly be equal.

Definition 8.30. Suppose V is a finite-dimensional inner product space. An operator $T \in \mathcal{L}(V)$ is said to be self-adjoint if $T = T^*$; that is, for all $v, w \in V$,

$$\langle T(v), w \rangle = \langle v, T(w) \rangle .$$

What do the matrix of such an operator look like? Well, $T = T^*$, so $\mathcal{M}(T) = \mathcal{M}(T^*) = \overline{\mathcal{M}(T)}^\top$ under any **orthonormal** basis (recall why this may not hold if the basis is not orthonormal!). In other words, the matrix of a self-adjoint operator is conjugate symmetric under an orthonormal basis. The converse turns out to hold as well from some calculation.

Proposition 8.31. Suppose V is a finite-dimensional inner product space and let $e_1, \cdots, e_n$ be an orthonormal basis of V. Then, for any $T \in \mathcal{L}(V)$, T is self-adjoint if and only if $\mathcal{M}(T)$ is conjugate symmetric.

Proof. Both directions are verified by direct calculation.

"If" direction. Suppose $\mathcal{M}(T) = \overline{\mathcal{M}(T)}^\top = \mathcal{M}(T^*)$ (Proposition 8.29). Then, $T = T^*$.

"Only if" direction. Suppose now that $T = T^*$. Then, $\mathcal{M}(T) = \mathcal{M}(T^*) = \overline{\mathcal{M}(T)}^\top$. □

Some very useful results follow immediately. The first one we present is the generalization of the following fact from MATH 220: the eigenvalues of conjugate symmetric matrices are real.

Proposition 8.32. Let $T \in \mathcal{L}(V)$ be a self-adjoint operator on a finite-dimensional inner product space V. Then, $E(T) \subset \mathbb{R}$.

Proof. Let $\lambda \in E(T)$ be an eigenvalue of T and fix $v \in E(T, \lambda) \backslash \{0\}$. Then,

$$\lambda \cdot \|v\|^2 = \langle T(v), v \rangle = \langle v, T(v) \rangle = \overline{\lambda} \cdot \|v\|^2 ,$$

so dividing both sides by $\|v\|^2 \neq 0$ gives $\lambda = \overline{\lambda}$, and hence $\lambda \in \mathbb{R}$. □

Some useful results from [1] are also presented as follows.

Proposition 8.33. Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$. Then, $T = 0$ if and only if $\langle T(v), v \rangle = 0$ for all $v \in V$.

Proof. The only if direction is obvious. Now suppose $\langle T(v), v \rangle = 0$ for any $v \in V$. Then, for all $u, w \in V$,

$$\langle T(u), w \rangle = \frac{\langle T(u+w), u+w \rangle - \langle T(u-w), u-w \rangle}{4} + \frac{\langle T(u+iw), u+iw \rangle - \langle T(u-iw), u-iw \rangle}{4} i = 0,$$

where the first equality can be seen by expanding the right hand side. The proof is finished. □

Note that the result above is clearly false for real inner product spaces. Consider, for example, the 90° counterclockwise rotation operator on $\mathbb{R}^2$. Every vector is perpendicular to itself rotated 90°, but this operator is not zero. While we can rotate in the real case to get this perpendicularity, this won't hold in the complex case: multiplication by an imaginary number reaches the rotation, so that perpendicularity cannot possibly hold.

While this fact seems out of nowhere, it is significant in the following way: we can simplify the definition of self-adjoint operators by changing v, w to just v:

Corollary 8.34. Suppose V is a complex inner product space and $T \in \mathcal{L}(V)$. Then, T is self-adjoint if and only if $\langle T(v), v \rangle = \langle v, T(v) \rangle$ for all $v \in V$. In other words, T is self-adjoint if and only if $\langle T(v), v \rangle \in \mathbb{R}$.

Proof. If T is self-adjoint, then taking $v = w$ in particular in the definition gives $\langle T(v), v \rangle = \langle v, T(v) \rangle$ for all $v \in V$. We now show the other direction, by supposing $\langle T(v), v \rangle = \langle v, T(v) \rangle$ for an arbitrary $v \in V$. Then,

$$\langle (T - T^*)(v), v \rangle = \langle T(v), v \rangle - \langle T^*(v), v \rangle = \langle v, T(v) \rangle - \langle v, T(v) \rangle = 0,$$

which implies $T - T^* = 0$ by the previous Proposition. Then, $T = T^*$, and T is self-adjoint. □

We end with another result about the perpendicularity, by replacing the condition of complex inner product space with the imposition that T be self-adjoint. Unfortunately, this proof also uses an equality that isn't quite familiar to us.

Proposition 8.35. Suppose V is a finite-dimensional inner product space and T is self-adjoint. Then, $T = 0$ if and only if $\langle T(v), v\rangle = 0$ for all $v \in V$.

Proof. We have already shown this is true for a complex inner product space. Now suppose V is a real inner product space. If $T = 0$, then the other follows immediately. We therefore show the other direction by supposing $\langle T(v), v\rangle = 0$ for an arbitrary $v \in V$. Then, for any $u, w \in V$,

$$\langle T(u), w\rangle = \frac{\langle T(u+w), u+w\rangle - \langle T(u-w), u-w\rangle}{4} = 0.$$

To show the first equality, we expand the right hand side:

$$\begin{aligned}
&\frac{\langle T(u+w), u+w\rangle - \langle T(u-w), u-w\rangle}{4} \\
&= \frac{\langle T(u), u\rangle + \langle T(u), w\rangle + \langle T(w), u\rangle + \langle T(w), w\rangle}{4} - \frac{\langle T(u), u\rangle - \langle T(u), w\rangle - \langle T(w), u\rangle + \langle T(w), w\rangle}{4} \\
&= \frac{\langle T(u), w\rangle + \langle T(w), u\rangle}{2} \\
&= \frac{\langle T(u), w\rangle + \langle w, T(u)\rangle}{2} \\
&= \frac{\langle T(u), w\rangle + \langle T(u), w\rangle}{2} \\
&= \langle T(u), w\rangle .
\end{aligned}$$

The proof is complete. □

Finally, we'll go back to the concept of orthogonal projection a bit. It turns out that every orthogonal projection is self-adjoint.

Proposition 8.36. Suppose V is a finite-dimensional inner product space and let U be a subspace. Then, P_U is self-adjoint.

Proof. Suppose $v, w \in V$ and fix an orthonormal basis $e_1, \cdots, e_k$ of U. Then, with all sums from $i = 1$ to k,

$$\begin{aligned}
\langle P_U(v), w\rangle &= \left\langle \sum_i \langle v, e_i\rangle \cdot e_i, w \right\rangle \\
&= \sum_i \langle v, e_i\rangle \cdot \langle e_i, w\rangle \\
&= \sum_i \overline{\langle w, e_i\rangle} \cdot \langle v, e_i\rangle \\
&= \left\langle v, \sum_i \langle w, e_i\rangle \cdot e_i \right\rangle \\
&= \langle v, P_u(w)\rangle .
\end{aligned}$$

The proof is complete. □

What's more, there is a converse to this: any idempotent, self-adjoint linear operator $T \in \mathcal{L}(V)$ is an orthogonal projection—specifically, onto its range. As a reminder, T is idempotent iff $T = T^2$.

Proposition 8.37. Suppose V is a finite-dimensional inner product space. Then, $T = P_U$ for some subspace U of V if and only if T is idempotent and self-adjoint. In this case, $U = \operatorname{range} T$.

Proof. One direction has already been shown. We now suppose $T = T^2$ and $T = T^*$. Fix an orthonormal basis $e_1, \cdots, e_k$ of $U := \operatorname{range} T$. Note that $T(e_i) = e_i$ for all $i \in \{1, \cdots, k\}$: because $e_i \in \operatorname{range} T$, fix $v_i \in V$ such that $T(v_i) = e_i$; then, $T(e_i) = T^2(v_i) = T(v_i) = e_i$. Hence,

$$\langle T(v), e_i \rangle = \langle v, T(e_i) \rangle = \langle v, e_i \rangle .$$

Because the coordinates of $T(v)$ and $P_U(v)$ agree under the orthonormal basis, $P_U = T$. □

8.6 Normal Operators

An important class of operators that we didn't talk about in depth in MATH 220 is the class of normal operators. In fact, as we will see later, in the complex case, this is the precisely the collection of all linear operators diagonalizable under an orthonormal basis.

Definition 8.38. Suppose V is a finite-dimensional inner product space. T is said to be normal if $T^*T = TT^*$.

A useful characterization of the normality is given by the norm.

Proposition 8.39. Suppose V is a finite-dimensional inner product space and $T \in \mathcal{L}(V)$. Then, T is normal if and only if $\|T(v)\| = \|T^*(v)\|$ for all $v \in V$.

Proof. Note that $T^*T - TT^*$ is a self-adjoint operator, because $(T^*T - TT^*)^* = (T^*T)^* - (TT^*)^* = T^*T - TT^*$. Then, Proposition 8.35 states $T^*T - TT^* = 0$ if and only if $\langle (T^*T - TT^*)(v), v \rangle = 0$ for all $v \in V$. Therefore,

$$\begin{aligned}
T \text{ is normal} \quad &\iff \quad T^*T - TT^* = 0 \\
&\iff \quad \langle (T^*T - TT^*)(v), v \rangle = 0 && \text{for all } v \in V \\
&\iff \quad \langle (T^*T)(v), v \rangle = \langle (TT^*)(v), v \rangle && \text{for all } v \in V \\
&\iff \quad \langle T(v), T(v) \rangle = \langle T^*(v), T^*(v) \rangle && \text{for all } v \in V \\
&\iff \quad \|T(v)\| = \|T^*(v)\| && \text{for all } v \in V.
\end{aligned}$$

The proof is complete. □

Our first big result will be about the eigenvalues and eigenvectors.

Theorem 8.40. *Suppose V is a finite-dimensional inner product space and let $T \in \mathcal{L}(V)$ be normal. If v is a λ-eigenvector of T, then v is also a $\overline{\lambda}$-eigenvector of T^*.*

Proof. First, for any eigenvalue $\lambda \in E(T)$, note that $T - \lambda \cdot I$ is normal too. Indeed,

$$\begin{aligned}
(T - \lambda \cdot I)^*(T - \lambda \cdot I) &= (T^* - \overline{\lambda} \cdot I)(T - \lambda \cdot I) = T^*T - \lambda \cdot T^* - \overline{\lambda} \cdot T + |\lambda|^2 \cdot I \\
(T - \lambda \cdot I)(T - \lambda \cdot I)^* &= (T - \lambda \cdot I)(T^* - \overline{\lambda} \cdot I) = TT^* - \lambda \cdot T^* - \overline{\lambda} \cdot T + |\lambda|^2 \cdot I.
\end{aligned}$$

Because $T^*T = TT^*$, the left hand sides agree as well.

Thus, for any eigenvector $v \in E(T, \lambda) \backslash \{0\}$, we have

$$\begin{aligned}
0 = \|T(v) - \lambda \cdot v\|^2 &= \langle (T - \lambda \cdot I)(v), (T - \lambda \cdot I)(v) \rangle \\
&= \langle v, (T - \lambda \cdot I)^*(T - \lambda \cdot I)(v) \rangle \\
&= \langle v, (T - \lambda \cdot I)(T - \lambda \cdot I)^*(v) \rangle \\
&= \langle (T - \lambda \cdot I)^*(v), (T - \lambda \cdot I)^*(v) \rangle \\
&= \|(T - \lambda \cdot I)^*(v)\|^2 .
\end{aligned}$$

Then, $(T^* - \overline{\lambda} \cdot I)(v) = 0$, so $v \in E(T, \overline{\lambda}) \backslash \{0\}$. □

Of course, this also goes both ways:

Corollary 8.41. Suppose V is a finite-dimensional inner product space and let $T \in \mathcal{L}(V)$ be normal. Then, $E(T^*) = \overline{E(T)}$ and $E(T,\lambda) = E(T^*, \overline{\lambda})$ for all $\lambda \in E(T)$.

Proof. We have shown inclusion of both in one direction. Now apply the Proposition to T^* and note $(T^*)^* = T$ to show inclusion in the other direction. □

Our last result is an extension of Proposition 6.5. For normal operators, eigenvectors corresponding to distinct eigenvalues are not just independent; they are orthogonal.

Proposition 8.42. Suppose $T \in \mathcal{L}(V)$ is a normal operator over a finite-dimensional inner product space V. Then, eigenvectors corresponding to distinct eigenvalues are orthogonal.

Proof. Let $v_1 \in E(T,\lambda_1)\backslash\{0\}$ be a λ_1-eigenvector and $v_1 \in E(T,\lambda_1)\backslash\{0\}$ a λ_2-eigenvector. Then, applying Proposition 8.40, we have

$$\begin{aligned}(\lambda_1 - \lambda_2)\langle v_1, v_2\rangle &= \langle \lambda_1 \cdot v_1, v_2\rangle - \left\langle v_1, \overline{\lambda}_2 \cdot v_2\right\rangle \\ &= \langle T(v_1), v_2\rangle - \langle v_1, T^*(v_2)\rangle \\ &= 0.\end{aligned}$$

Because $\lambda_1 \neq \lambda_2$, this implies $\langle v_1, v_2\rangle = 0$, as desired. □

8.7 Spectral Theorems

We've already seen before some equivalent conditions of diagonalizability: we need a basis of eigenvectors, or equivalently, the direct sum of all eigenspaces is the vector space. In inner product spaces, we have more ways to classify operators, like self-adjoint and normal ones. How do these definitions in an inner product space interact with the concept of diagonalizability?

First, we can add one additional qualification to our previous definition of diagonalizability: we will look for operators that are diagonalizable under an orthonormal basis.

Definition 8.43. Suppose V is a finite-dimensional inner product space. An operator $T \in \mathcal{L}(V)$ is said to be orthogonally diagonalizable if $\mathcal{M}(T)$ is diagonal under some orthonormal basis.

Here, we can replace the phrasing "orthonormal basis" at the end with "orthogonal basis." If $\mathcal{M}(T)$ is diagonal under an orthogonal basis, then $\mathcal{M}(T)$ is clearly the same matrix in the orthonormal basis from normalizing the old one.

But, this additional assumption that the basis is orthogonal/orthonormal is not trivial: not all diagonalizable operators are orthogonally diagonalizable. For example, if $T \in \mathcal{L}(\mathbb{R}^2)$ is the operator that doesn't change $(1,0)$ but doubles $(1,1)$, it is clearly diagonalizable under the basis formed by these two vectors. However, there's no way to get rid of the 45° angle between the two vectors: the matrix is not orthogonally diagonalizable in the inner product space.

How might this definition be useful? Let's take a look from a computational standpoint as a computer scientist. Say $T \in \mathcal{L}(\mathbb{R}^n)$ is diagonalizable as

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

under some basis $v_1, \cdots, v_n$ and we want to compute $T^k(v)$ for a large k. If we have the coordinates $v = c_1 \cdot v_1 + \cdots + c_n \cdot v_n$, then the actual computation is simply $T^k(v) = c_1\lambda_1^k \cdot v_1 + \cdots + c_1\lambda_n^k \cdot v_n$ which is $O(n)$. Even without an orthonormal basis, we have a significant speedup against the $O(kn^2)$ naive calculation.

But, we still need

$$\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}^{-1} \cdot v$$

to get the coefficients with a regular basis. However, if it is an orthonormal basis, we can simply do

$$c_i = \langle v, v_i \rangle .$$

Of course, caching the inverse matrix in advance means both take $O(n^2)$ in practice, but the latter is nonetheless much preferable in implementation.

Okay, now we know why orthogonally diagonalizable matrices are nice, and we also established not all diagonalizable matrices are orthogonally diagonalizable. So what are these matrices exactly? It turns out they're precisely self-adjoint and normal operators in the real and complex cases, respectively. These two theoretically beautiful results are known as spectral theorems.

We will first tackle the complex case, for which we have enough machinery at this point.

Theorem 8.44 (Complex Spectral Theorem)**.** *Suppose V is a finite-dimensional complex inner product space and $T \in \mathcal{L}(V)$. Then, T is orthogonally diagonalizable if and only if T is normal.*

Note that the only operator on a 0-dimensional inner product space is the identity/zero operator, which is trivially normal and orthogonally diagonalizable. The following proof addresses the cases with $\dim V \in \mathbb{N}_+$.

Proof. The "only if" direction is direct. We perform induction on $\dim V$ for the other direction.

"If" direction. Let $T \in \mathcal{L}(V)$ be normal. The base case is trivial: all operators in a 1-dimensional inner product space are both normal and orthogonally diagonalizable. We now suppose inductively that $n := \dim V > 1$ and all normal operators on an $(n-1)$-dimensional real inner product space are orthogonally diagonalizable. Fix an eigenvalue $\lambda_1 \in E(T)$ of T by Theorem 6.10 as well as an associated eigenvector $v_1 \in E(T, \lambda_1) \backslash \{0\}$. Let $e_1 := v_1 / \|v_1\|$, which is also a λ_1-eigenvector. Consider the $(n-1)$-dimensional orthogonal complement $W := \operatorname{span}(e_1)^\perp$, which we claim is T-invariant. Indeed, for an arbitrary $w \in W$, $\langle T(w), e_1 \rangle = \langle w, T^*(e_1) \rangle = \left\langle w, \overline{\lambda}_1 \cdot e_1 \right\rangle = \lambda_1 \cdot 0 = 0$ by Proposition 8.40. Hence, $T|_W \in \mathcal{L}(W)$ on the $(n-1)$-dimensional W is orthogonally diagonalizable under some orthonormal basis $e_2, \cdots, e_n$ of W as

$$\mathcal{M}(T|_W) = \begin{pmatrix} \lambda_2 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}.$$

Then, $T(e_i) = \lambda_i \cdot e_i$ for all $i \in \{1, \cdots, n\}$, where $e_1, \cdots, e_n$ is an orthonormal basis of V. Then, T is orthogonally diagonalizable under $e_1, \cdots, e_n$.

"Only if" direction. Let $\mathcal{M}(T)$ be diagonal with diagonal entries $\lambda_1, \cdots, \lambda_n$ under an orthonormal basis $e_1, \cdots, e_n$. Then, $\mathcal{M}(T^*) = \overline{\mathcal{M}(T)}^\top$ is also diagonal. Because diagonal matrices commute, $\mathcal{M}(TT^*) = \mathcal{M}(T^*T)$, and this implies $TT^* = T^*T$. □

There's a pitfall: even though $\mathcal{M}(T)$ is diagonal, this is not necessarily conjugate symmetric! Note that the diagonal may be complex and not real, in which case T cannot be self-adjoint. Indeed, if all eigenvalues of a normal operator are real, the above argument shows that the normal operator is further self-adjoint.

In the real case, things are a bit more subtle. It turns out that not all normal matrices are orthogonally diagonalizable. Consider, for example, $T \in \mathcal{L}(\mathbb{R}^3)$ defined by $T(x, y, z) := (x - z, y, x + z)$ or

$$\mathcal{M}(T) = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Whether real or complex, T is normal but not self-adjoint. Considered as a complex operator, T has 3 distinct eigenvalues $1, 1+i, 1-i$, but that means T only has one eigenvalue as a real operator, and there's no way to get an eigenbasis that is orthonormal.

The stronger condition we need is that T must be self-adjoint. A real operator may not have an eigenvalue, which is precisely where the proof in the complex case fails here. But a self-adjoint operator always has an eigenvalue. Intuitively, complexifying it gives a complex operator that will have an eigenvalue, which must be real.

Lemma 8.45. Let V be a finite-dimensional real inner product space and suppose $T \in \mathcal{L}(V)$ is self-adjoint. Then, T has an eigenvalue.

Proof. Fix an orthonormal basis $e_1, \cdots, e_n$ of V. Then, $\mathcal{M}(T)$ is conjugate symmetric.

We make $V_{\mathbb{C}}$ an inner product space by endowing the inner product

$$\langle v, w\rangle_{V_{\mathbb{C}}} := x_1\overline{y}_1 + \cdots + x_n\overline{y}_n$$

where $v = x_1 \cdot e_1 + \cdots + x_n \cdot e_n \in V_{\mathbb{C}}$ and $w = y_1 \cdot e_1 + \cdots + y_n \cdot e_n \in V_{\mathbb{C}}$. The axioms are satisfied, which guarantees the validity of the inner product. Note that $e_1, \cdots, e_n$ remains an orthonormal basis in $V_{\mathbb{C}}$.

We claim that $T_{\mathbb{C}}$ is self-adjoint. Indeed, because $\mathcal{M}(T) = \mathcal{M}(T_{\mathbb{C}})$ by Proposition 7.21, $\mathcal{M}(T_{\mathbb{C}})$ is also conjugate symmetric, and Proposition 8.31 implies that $T_{\mathbb{C}}$ is also self-adjoint. Fix an eigenvalue $\lambda \in \mathbb{C}$ of $T_{\mathbb{C}}$. Then, λ must be real by Proposition 8.32.

Because $\chi_T = \chi_{T_{\mathbb{C}}}$, we conclude that $\lambda \in \mathbb{R}$ must be an eigenvalue of T. □

The real spectral theorem now follows the same logic.

Theorem 8.46 (Real Spectral Theorem)**.** *Suppose V is a finite-dimensional real inner product space and $T \in \mathcal{L}(V)$. Then, T is orthogonally diagonalizable if and only if T is self-adjoint.*

We assume $\dim V \in \mathbb{N}_+$ as well.

Proof. The "only if" direction is direct. We perform induction on $\dim V$ for the other direction.

"If" direction. Let $T \in \mathcal{L}(V)$ be self-adjoint. The base case is trivial: all operators in a 1-dimensional inner product space are both self-adjoint and orthogonally diagonalizable. We now suppose inductively that $n := \dim V > 1$ and all self-adjoint operators on an $(n-1)$-dimensional real inner product space are orthogonally diagonalizable. Fix an eigenvalue $\lambda_1 \in E(T)$ of T by the Lemma above as well as an associated eigenvector $v_1 \in E(T, \lambda_1)\backslash\{0\}$. Let $e_1 := v_1/\|v_1\|$, which is also a λ_1-eigenvector. Consider the $(n-1)$-dimensional orthogonal complement $W := \operatorname{span}(e_1)^{\perp}$, which we claim is T-invariant. Indeed, for an arbitrary $w \in W$, $\langle T(w), e_1\rangle = \langle w, T^*(e_1)\rangle = \left\langle w, \overline{\lambda}_1 \cdot e_1\right\rangle = \lambda_1 \cdot 0 = 0$ by Proposition 8.40. Hence, $T|_W \in \mathcal{L}(W)$ on the $(n-1)$-dimensional W is orthogonally diagonalizable under some orthonormal basis $e_2, \cdots, e_n$ of W as

$$\mathcal{M}(T|_W) = \begin{pmatrix} \lambda_2 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}.$$

Then, $T(e_i) = \lambda_i \cdot e_i$ for all $i \in \{1, \cdots, n\}$, where $e_1, \cdots, e_n$ is an orthonormal basis of V. Then, T is orthogonally diagonalizable under $e_1, \cdots, e_n$.

"Only if" direction. Let $\mathcal{M}(T)$ be diagonal with diagonal entries $\lambda_1, \cdots, \lambda_n$ under an orthonormal basis $e_1, \cdots, e_n$. A real diagonal matrix is symmetric, so Proposition 8.31 implies that T is self-adjoint. □

8.8 Isometries

Prototypical Example. The operator on $\mathbb{R}^2$ that first rotates 90° and then flips w.r.t. the y-axis.

We now consider another family of linear operators over an inner product space—the ones that preserve the length of any vector. For example, in $\mathbb{R}^2$, any rotation is isometry, and so is any reflection w.r.t. a line. We now make this idea more precise, and look into some of their properties.

Definition 8.47. Suppose V is an inner product space. An operator $S \in \mathcal{L}(V)$ is said to be an <u>isometry</u> if $\|S(v)\| = \|v\|$ for all $v \in V$.

It turns out that these are also precisely the operators that preserve the inner product, a seemingly much stronger/harder-to-meet condition.

Proposition 8.48. Suppose V is an inner product space and $S \in \mathcal{L}(V)$. Then, S is an isometry if and only if $\langle S(v), S(w)\rangle = \langle v, w\rangle$ for all $v, w \in V$.

Proof. The "if" direction is obvious by substituting $w = v$. We show the "only if" direction through the polarization inequality, supposing S is an isometry.

If the field underlying V is $\mathbb{R}$, then for all $v, w \in V$, we have through the polarization identity

$$\begin{aligned}\langle S(v), S(w)\rangle &= \frac{1}{4}(\|S(v)+S(w)\|^2 - \|S(v)-S(w)\|^2)\\ &= \frac{1}{4}(\|S(v+w)\|^2 - \|S(v-w)\|^2)\\ &= \frac{1}{4}(\|v+w\|^2 - \|v-w\|^2)\\ &= \langle v, w\rangle .\end{aligned}$$

If the field underlying V is $\mathbb{R}$, then for all $v, w \in V$, we have through the polarization identity

$$\begin{aligned}\langle S(v), S(w)\rangle &= \frac{1}{4}(\|S(v)+S(w)\|^2 - \|S(v)-S(w)\|^2 + i\,\|S(v)+iS(w)\|^2 - i\,\|S(v)-iS(w)\|^2)\\ &= \frac{1}{4}(\|S(v+w)\|^2 - \|S(v-w)\|^2 + i\,\|S(v+iw)\|^2 - i\,\|S(v-iw)\|^2)\\ &= \frac{1}{4}(\|v+w\|^2 - \|v-w\|^2 + i\,\|v+iw\|^2 - i\,\|v-iw\|^2)\\ &= \langle v, w\rangle .\end{aligned}$$

The proof is now complete. □

We know that an eigenvalue stretches associated eigenvectors by a scalar factor of that eigenvalue. It is then intuitive that we can only stretch by "1," or more precisely, by scalars whose absolute value is 1. Indeed, we have the following result.

Proposition 8.49. Suppose V is an inner product space and $S \in \mathcal{L}(V)$ is an isometry. Then, any eigenvalue of S has modulus 1.

Proof. Suppose λ is an eigenvalue of S. Let $v \in E(S, \lambda)$ be an associated eigenvector. Then,

$$\|v\| = \|S(v)\| = \|\lambda \cdot v\| = |\lambda| \cdot \|v\| .$$

Because $\|v\| \neq 0$, dividing by the norm from both sides yields $|\lambda| = 1$. □

The matrix of such operators are known by the name of orthogonal matrices in the real case, and unitary matrices in the complex case. A result from MATH 220 is that its inverse equals its conjugate transpose. Indeed, we have the following:

Proposition 8.50. Suppose V is a finite-dimensional inner product space and $S \in \mathcal{L}(V)$. Then, S is an isometry if and only if $S^*S = I$. If this holds, then further $S^*S = SS^* = I$.

Proof. Both directions are straightforward calculations.

"If" direction. Suppose $S^*S = I$. Then, for all $v, w \in V$,

$$\langle S(v), S(w) \rangle = \langle (S^*S)(v), w \rangle = \langle v, w \rangle ,$$

so Proposition 8.48 implies that S is an isometry.

"Only if" direction. Let S be an isometry. Fix an orthonormal basis $e_1, \cdots, e_n$ of V. By Proposition 8.48,

$$\mathcal{M}(S^*S)_{j,i} = \left\langle (S^*S)(e_i), e_j \right\rangle = \left\langle Se_i, Se_j \right\rangle = \left\langle e_i, e_j \right\rangle = \delta_{i,j} = \mathcal{M}(I)_{j,i},$$

so $S^*S = I$.

Now suppose both are simultaneously true. Because the composition of S^* and S is injective, both S^* and S must be injective, and hence invertible (Proposition 4.20). The uniqueness of the inverse then implies $S^*S = SS^* = I$. □

Another related idea is how an orthonormal basis is mapped under an isometry: an isometry preserves orthonormal bases. We actually have something even stronger:

Proposition 8.51. Suppose V is an inner product space and let $S \in \mathcal{L}(V)$. Then, S is an isometry if and only if $Se_1, \cdots, Se_n$ is also an orthonormal basis of V for *some* orthonormal basis $e_1, \cdots, e_n$ of V.

Proof. We make extensive use of the Pythagorean theorem (Theorem 8.4) to bridge the gap between norms and the orthonormality of the bases.

"If" direction. Suppose $Se_1, \cdots, Se_n$ is an orthonormal basis of V for *some* orthonormal basis $e_1, \cdots, e_n$ of V. Then, the Pythagorean theorem implies

$$\|S(v)\| = \left\| S\left(\sum_{i=1}^{n} \langle v, e_i \rangle \cdot e_i \right) \right\| = \left\| \sum_{i=1}^{n} \langle v, e_i \rangle \cdot S(e_i) \right\| = \sum_{i=1}^{n} |\langle v, e_i \rangle| .$$

At the same time,

$$\|v\| = \left\| \sum_{i=1}^{n} \langle v, e_i \rangle \cdot e_i \right\| = \sum_{i=1}^{n} |\langle v, e_i \rangle|$$

also, so S is an isometry by definition.

"Only if" direction. Now suppose instead that S is an isometry. Then,

$$\left\langle S(e_i), S(e_j) \right\rangle = \left\langle (S^*S)(e_i), e_j \right\rangle = \left\langle e_i, e_j \right\rangle = \delta_{i,j},$$

so $S(e_1), \cdots, S(e_n)$ is an orthonormal list long enough to be a basis. □

References

[1] Sheldon Axler. Linear algebra done right. Springer, 2015.