18.700: Linear Algebra

Jack Liu

Fall 2022

Contents

1 $\mathbb{R}^n$ and Abstract Addition and Scalar Multiplication

1.1 Introduction

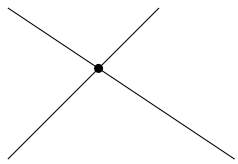
The lecturer is **Zhiwei Yun**. The notes are taken by **Jack Liu** and the note-taking is supervised by **Ashay Athalye**. The text used in this class is the 3rd edition of **Linear Algebra Done Right**, by Sheldon Axler.

This class is different from a linear algebra class you might have taken in the past. Most linear algebra classes emphasize matrices and solving linear equations. However, this class will focus on proofs and the theoretical side of linear algebra.

A typical problem in linear algebra involves solving a system of linear equations:

$$\begin{cases} 2x + 3y = 1 \\ x - \ \ y = 2. \end{cases}$$

We can also interpret these equations as lines on a plane, with the solution being the intersection of the two lines.



Thus, we can describe linear algebra both algebraically (by solving linear equations) and geometrically (by finding the intersection of two "flat" shapes or finding the distance from a point to a "flat" shape).[1]

The central objects of this course are *vector spaces*. We will spend the first several lectures investigating many aspects of vector spaces. Afterwards, we will study *linear maps*, which will relate vector spaces to each other. Vector spaces and linear maps are examples of *abstract* objects. The opposite of abstract objects are *concrete* objects, such as a vector (x_1, x_2, x_3) or the matrix

$$\begin{pmatrix} 2 & 0 & 4 \\ 0 & 7 & 8 \\ 1 & 0 & 1 \end{pmatrix}.$$

These objects are concrete in the sense that they can be represented by numbers, while abstract objects cannot. Throughout this course, we will develop a bridge between abstract objects and concrete objects, and show how these objects actually represent the same idea in mathematics. While we will touch on both types of objects, we will mostly focus on the abstract objects of vector spaces and linear maps.

Finally, a large part of this class is understanding and writing proofs. Each lecture will present at least one proof, and the homework and exams will have multiple proofs for you to derive.

1.2 $\mathbb{R}^n$

Now, we will introduce the two most basic examples of a vector space.

Definition 1.1 ($\mathbb{R}^n$)
$\mathbb{R}^n$ is the set of all lists of length n of elements of $\mathbb{R}$:

$$\mathbb{R}^n = \{(x_1, \ldots, x_n) \mid x_i \in \mathbb{R} \text{ for } i = 1, \ldots, n\}.$$

The curly braces around the definition of $\mathbb{R}^n$ signify that $\mathbb{R}^n$ is a set. The $(x_1, \ldots, x_n)$ gives the general form of an element of $\mathbb{R}^n$ as a list of length n, while the $x_i \in \mathbb{R}$ denotes that each element of the list is a real number.

In addition, here is a quick review on set notation.

[1] The term "flat" is not well-defined, but it comes from the fact that the equations for these shapes must be linear. Thus, lines and planes are "flat" shapes, while circles and ellipses are not.

Example 1.2
Suppose S is a set.

- $a \in S$: a is an element of S
- $S \subset S'$: S is a subset of S'
- $\forall$: for all
- $\exists$: there exists
- $\exists!$: there exists a unique.

For instance, the notation

$$\exists!\, x \in \mathbb{R} \text{ s.t. } 2x = 1$$

means "there exists a unique element x in $\mathbb{R}$ such that $2x = 1$."

$\mathbb{R}^n$ is a relatively simple example of a vector space. First, let's investigate some operations on $\mathbb{R}^n$.

Definition 1.3 (Operations on $\mathbb{R}^n$)
Suppose $(x_1, \dots, x_n), (\eta_1, \dots, y_n) \in \mathbb{R}$ and $a \in \mathbb{R}$. The, we define the following definitions:

1. addition: $(x_1, \dots, x_n) + (y_1, \dots, y_n) = (x_1 + y_1, \dots, x_n + y_n)$
2. multiplication: $(x_1, \dots, x_n) \cdot (y_1, \dots, y_n) = (x_1 y_1, \dots, x_n y_n)$
3. scalar multiplication: $a \cdot (x_1, \dots, x_n) = (ax_1, \dots, ax_n)$.

Note that these are not the only operations that could be defined on $\mathbb{R}^n$ (for instance, one could easily define subtraction).

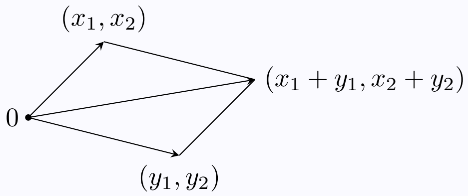
Student Question. *What is the geometric interpretation of the multiplication of two elements of $\mathbb{R}^n$?*

Answer. *In fact, there is no natural geometric interpretation of multiplication. Thus, for most of this course, we will completely forget about multiplication, and only focus on addition and scalar multiplication.*[2]

On the other hand, addition and scalar multiplication both have natural geometric meanings.

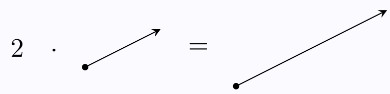
Example 1.4
Suppose $(x_1, x_2), (y_1, y_2) \in \mathbb{R}^2$. Then, $(x_1 + y_1, x_2 + y_2)$ represents the fourth vertex of a parallelogram with vertices $(x_1, y_1), (x_2, y_2)$, and the origin:

addition:



Additionally, scalar multiplication represents a scaling the length of a vector by some factor:

scalar multiplication:



Student Question. *Do the geometrical meaning of addition and scalar multiplication extend to higher dimensions of $\mathbb{R}^n$?*

Answer. *Yes! For instance, the addition of three vectors in $\mathbb{R}^3$ forms a parallelepiped, analogous to the parallelogram in $\mathbb{R}^2$.*

[2] We will see some operations that look like multiplication when we cover inner product spaces, which will be much later in this course.

There also exists a special vector in $\mathbb{R}^n$, known as the zero vector.

Definition 1.5 ($\vec{0}$)
Let $\vec{0}$ denote the element in $\mathbb{R}^n$ defined by

$$\vec{0} = (0, \ldots, 0).$$

For all $x = (x_1, \ldots, x_n) \in \mathbb{R}$, $\vec{0}$ satisfies

$$\vec{0} + x = x + \vec{0} = x.$$

Thus, $\vec{0}$ is an additive identity for $\mathbb{R}^n$.

The following result gives more properties of addition.

Theorem 1.6 (Properties of addition in $\mathbb{R}^n$)
Suppose $x, y, z \in \mathbb{R}^n$. Addition satisfies the following properties:

- commutativity: $x + y = y + x$
- associativity: $(x + y) + z = x + (y + z)$
- additive inverse: $-x = (-x_1, \ldots, -x_n)$.

In particular, we can think of the additive inverse of x in two ways. Using scalar multiplication, we can express

$$-x = (-1)x.$$

Using addition, we can express

$$x + (-x) = \vec{0}.$$

The final property of $\mathbb{R}^n$ that we will discuss in this section is as follows.

Theorem 1.7 (distributive property)
Suppose $x, y \in \mathbb{R}^n$ and $a, b \in \mathbb{R}$. Then,

$$a(x + y) = ax + ay$$

and

$$(a + b)x = ax + bx.$$

In summary, $\mathbb{R}^n$ is a set with two operations: addition and scalar multiplication. While many of these properties of $\mathbb{R}^n$ may seem quite basic, we will carry over many of these properties in general vector spaces.

1.3 Examples of Vector Spaces

Our discussion of $\mathbb{R}^n$ in the previous section gives us some intuition on what a vector space is. As a rough definition, a vector space V is a set with two operations, abstract addition and abstract scalar multiplication, that satisfies a list of properties. Thus, when we define a vector space, we have to give three pieces of structure: a set of vectors, abstract addition, and abstract scalar multiplication.

Let us consider more examples of vector spaces. Although we have not fully defined vector spaces yet, we can verify later that all of these examples are indeed vector spaces.

Example 1.8
Consider the set of complex numbers $\mathbb{C}$. We can express any complex number as $a + bi$ for $a, b \in \mathbb{R}$. Thus, $(a, b) \in \mathbb{R}^2$, so the set of elements $\mathbb{C}$ is essentially equivalent to the set of elements $\mathbb{R}^2$.

Addition in $\mathbb{C}$ also works analogously to $\mathbb{R}^2$, since addition works by combining the real parts and imaginary parts separately:

$$(a + bi) + (c + di) = (a + c) + (b + d)i.$$

However, multiplication in $\mathbb{C}$ is more complicated. We know that $\mathbb{R}^2$ has scalar multiplication defined by

$$t(a, b) = (ta, tb)$$

for $t \in \mathbb{R}$. However, multiplication in $\mathbb{C}$ is defined by

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i.$$

It follows that multiplication in $\mathbb{C}$ is more general than scalar multiplication in $\mathbb{R}$. Setting $a + bi = t \in \mathbb{R}$ (i.e. $b = 0$) gives

$$(t + 0i)(c + di) = tc + tdi,$$

which is analogous to scalar multiplication in $\mathbb{R}^2$.

Student Question. *For scalar multiplication, must is a scalar a in $\mathbb{R}$ or $\mathbb{C}$?*

Answer. *In $\mathbb{R}^n$, we are working with scalars in $\mathbb{R}$. Later, we will be working with both $\mathbb{R}$ and $\mathbb{C}$. In particular, when we introduce vector spaces, we will have to state whether V is a real vector space or a complex vector space, which will determine whether a is in $\mathbb{R}$ or $\mathbb{C}$. Therefore, we actually have to specify four things to define a vector space: a field $\mathbb{F}$ ($\mathbb{R}$ or $\mathbb{C}$), a set of vectors, addition, and scalar multiplication.*

In the example above, we saw that $\mathbb{C}$ and $\mathbb{R}^2$ have the same structure in terms of the set of vectors, addition, and scalar multiplication if you restrict the scalars to be real. Thus, $\mathbb{C}$ is a real vector space.

Example 1.9
Suppose S is an arbitrary set and $\mathbb{F} = \mathbb{R}$. Let

$$V = \{f : S \to \mathbb{R}\}.$$

We wish to define abstract addition and scalar multiplication to make V into a vector space.

For abstract addition, suppose $f, g \in V$. We define

$$(f + g)(x) = f(x) + g(x)$$

for all $x \in S$.

For abstract scalar multiplication, suppose $f \in V$ and $a \in \mathbb{R}$. Suppose that we define

$$(af)(x) = f(ax)$$

for all $x \in S$. However, note that S is an arbitrary set, so it could be possible that ax is not well-defined. For instance, S could be a set consisting of apples, so it would not make sense to multiply an apple by some scalar a. Instead, we define

$$(af)(x) = a \cdot f(x).$$

Note that $f(x) \in \mathbb{R}$, so it always possible to multiply $a \cdot f(x)$. Thus, this definition makes sense.

Example 1.10
Suppose $\mathbb{F} = \mathbb{R}$ and

$$V = \{f : [0,1] \to \mathbb{R}, \text{ continuous functions}\}.$$

Define abstract addition and scalar multiplication as they are defined in Example 1.9.

Note that this example is a special case of Example 1.9, where $S = [0,1]$ and we are restricting f to by continuous.

In the above two examples, we specified the four pieces of structure that make up a vector space (field, set of vectors, addition, and scalar multiplication). However, we did not check any properties of V, so we cannot say yet whether V is a vector space or not. Once we fully define vector spaces, we will be able to come back and verify that V is indeed a vector space in both examples.

Student Question. *Is it possible for two vector spaces to have the same $\mathbb{F}$ and V, but have di erent definitions of abstract addition and scalar multiplication?*

Answer. *Yes! Suppose $\mathbb{F} = \mathbb{R}$ and $V = \mathbb{R}$. Define abstract addition as the usual addition in $\mathbb{R}$. Now, we will accept the following fact.*

Fact 1.11
$\mathbb{R}$ has a lot of automorphisms.

An automorphism is a bijection $\varphi : \mathbb{R} \to \mathbb{R}$ which preserves the arithmetic operations

$$\varphi(x + y) = \varphi(x) + \varphi(y)$$

and

$$\varphi(xy) = \varphi(x)\varphi(y).$$

It is impossible to explicitly write down any such φ (besides the identity). However, in an abstract sense, there exist an uncountably infinite such φ.

Then, define abstract scalar multiplication as

$$a \cdot x = \varphi(a)x.$$

It turns out that this "weird" definition of V is indeed a vector space. Another example is $\mathbb{F} = \mathbb{C}$, $V = \mathbb{R}$, addition is the usual addition, and scalar multiplication is defined by $a \cdot x = \overline{a}x$.

This shows that for any given field and set of vectors, there can still be several di erent definitions of abstract addition and scalar multiplication to form a vector space.

Example 1.12
Suppose $\mathbb{F} = \mathbb{R}$ and

$$V = \{f : \mathbb{R} \to \mathbb{R}, \text{ differentiable functions} \mid f'(x) + x^2 f(x) = 0\}.$$

Define abstract addition and scalar multiplication as as they are defined in Example 1.9. For any $f, g \in V$, let us verify that $h = f + g$ satisfies $h' + x^2 h = 0$. It is clear that

$$h' + x^2 h = (f' + g') + x^2(f + g) = (f' + x^2 f) + (g' + x^2 g) = 0,$$

so $h \in V$. Furthermore, we can similarly verify that $h = af$ satisfies $h' + x^2 h = 0$ for any $a \in \mathbb{R}$. Therefore, addition and scalar multiplication are well-defined.

However, suppose

$$V = \{f : \mathbb{R} \to \mathbb{R}, \text{ differentiable functions} \mid f'(x) + x^2 f(x) = 1\}.$$

Then, for $f, g \in V$, we can compute that $h = f + g$ satisfies

$$h' + x^2 h = 2 \neq 1.$$

Thus, this definition of abstract addition does not define an operation on V.

Note that the set $V = \{f \mid f' + x^2 f = 1\}$ is a subset of the set $\{f \mid \mathbb{R} \to \mathbb{R}\}$, which we gave vector space structure in Example 1.9 (set $S = \mathbb{R}$). To show that V is a vector space with the same definition of abstract addition and scalar multiplication, we must show that the result of abstract addition or scalar multiplication of any elements of V is still in V, which turns out not to be the case in this example. This will lead to the notion of subspace, which we will discuss in a later lecture.

2 Vector Spaces and Subspaces

2.1 Review

Last time, we introduced one of the simplest vector spaces, $\mathbb{R}^n$, and discussed some of its properties. We also used the properties of $\mathbb{R}^n$ to motivate some other examples of vector spaces and gave a very rudimentary definition of what a vector space is.

2.2 Definition of Vector Space

Now, we will finally fully define vector spaces. Recall that $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

Definition 2.1 (vector space)
A **$\mathbb{F}$-vector space** (also called a vector space over $\mathbb{F}$) is a set V along with two operations

- addition, which maps each pair of elements $u, v \in V$ to an element $u + v \in V$,
- scalar multiplication, which maps each $a \in \mathbb{F}$ and each $u \in V$ to an element $av \in V$

satisfying the following properties:

- commutativity:
$$u + v = v + u \text{ for all } u, v \in V;$$
- associativity:
$$(u + v) + w = u + (v + w) \text{ for all } u, v, w \in V$$
$$(ab)v = a(bv) \text{ for all } a, b \in \mathbb{F},\ v \in V;$$
- additive identity:
$$\text{there exists an element } 0 \in V \text{ such that } 0 + v = v \text{ for all } v \in V;$$
- additive inverse:
$$\text{for every } v \in V, \text{ there exists } w \in V \text{ such that } v + w = \vec{0};$$
- multiplicative identity:
$$1v = v \text{ for all } v \in V;$$
- distributive properties:
$$a(u + v) = au + av \text{ for all } a \in \mathbb{F},\ u, v \in V$$
$$(a + b)v = av + bv \text{ for all } a, b \in \mathbb{F},\ v \in V.$$

Student Question. *Since we mentioned commutativity of addition, does there also exists commutativity of scalar multiplication, i.e. are av and va for $a \in \mathbb{F}$ and $v \in V$?*

Answer. *We always denote scalar multiplication by putting the scalar in front of the vector, i.e. av. We will never write va.*

Remark. *We will soon prove that if w is the additive inverse of v, then w is unique. In fact, $w = (-1)v$. Since w is unique, we will denote w by $-v$.*

Now, consider the following example of a vector space, known as the *function space*.

Example 2.2
Suppose S is a set and
$$V = \{f : S \to \mathbb{C}\}.$$
For $f, g \in V$, define addition as
$$(f + g)(x) = f(x) + g(x)$$
for all $x \in S$. This defines a new function $f + g : S \to \mathbb{C}$.

For $a \in \mathbb{C}$ and $f \in V$, define scalar multiplication as
$$(af)(x) = a \cdot f(x)$$
for all $x \in S$. Similarly, this defines a new function $af : S \to \mathbb{C}$.

Let us verify that V has an additive identity. Note that V must contain the zero function:
$$\vec{0}(x) = 0$$
for all $x \in S$. For any, $f \in V$,
$$(f + \vec{0})(x) = f(x) + \vec{0}(x) = f(x)$$
for all $x \in S$. Therefore, $f + \vec{0} = f$, so $\vec{0}$ is the additive identity in V. The verification that V satisfies all other properties in Definition 2.1 is left as an exercise.

The following example will introduce $\mathbb{C}^n$, which is analogous to $\mathbb{R}^n$ over complex numbers.

Example 2.3
Define
$$\mathbb{C}^n = \{(x_1, \ldots, x_n) \mid x_i \in \mathbb{C},\ i = 1, \ldots, n\}.$$
Furthermore, define V as in Example 2.2 with $S = \{1, \ldots, \}$. Then, there exists a bijection from V to $\mathbb{C}^n$ by setting $f(i) = x_i$ for all $i = 1, \ldots, n$. Thus, $\mathbb{C}^n$ is a vector space.

2.3 Properties of Vector Spaces

Now, we will discuss some consequences of the 6 basic properties in Definition 2.1.

Theorem 2.4
Suppose V is a vector space. Then, the additive identity in V is unique.

Proof. Suppose $\vec{0}$ and $\vec{0}'$ are both additive identities in V. We wish to show that $\vec{0} = \vec{0}'$. We know that $\vec{0} + v = v$ and $\vec{0}' + v = v$ for all $v \in V$. Then,
$$\vec{0} = \vec{0} + \vec{0}' = \vec{0}' + \vec{0} = \vec{0}',$$
as desired. □

The next result will show another uniqueness property.

Theorem 2.5
Suppose V is a vector space. Then, any element in V has a unique additive inverse.

Proof. Let $v \in V$. Suppose w and w' are both additive inverses of v. We know that $v + w = v + w' = \vec{0}$. Then,
$$w = \vec{0} + w = (w' + v) + w = w' + (v + w) = w' + \vec{0} = w',$$
as desired. □

Thus, since additive inverse is unique, we can wrote $-v$ to denote the additive inverse of v.

Theorem 2.6
Suppose V is a vector space $u, v, w \in V$. If $u + w = v + w$, then $u = v$.

Proof. Adding $-w$ to both sides gives the desired result. □

Theorem 2.7
Suppose V is a vector space. Let $a \in \mathbb{F}$ and $v \in V$. Then,

1. $a\vec{0} = \vec{0}$
2. $0v = \vec{0}$
3. $(-1)v = -v$.

Proof. To prove (1), we have

$$a\vec{0} = a(\vec{0} + \vec{0}) = a\vec{0} + a\vec{0}.$$

Adding $-(a\vec{0})$ to both sides gives $a\vec{0} = \vec{0}$.

To prove (2), we have

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

Adding $-(0v)$ to both sides gives $0v = 0$.

To prove (3), we have

$$v + (-1)v = 1v + (-1)v = (1 + (-1))v = 0v = \vec{0},$$

so $(-1)v$ is the additive inverse of v. □

2.4 Subspaces

In this section, we will introduce the notion of subspaces.

Definition 2.8
Suppose V is a vector space. A subset $U \subset V$ is called a **subspace** of V is it satisfies:

1. $\vec{0} \in U$
2. U is closed under addition (i.e. $u_1, u_2 \in U$ implies $u_1 + u_2 \in U$)
3. U is closed under scalar multiplication (i.e. $a \in \mathbb{F}$, $u \in U$ implies $au \in U$).

Consider the following example of a subspace.

Example 2.9
Suppose $V = \mathbb{C}^2 = \{(x_1, x_2) \mid x_1, x_2 \in \mathbb{C}\}$. Let

$$U_1 = \{(x_1, 0) \mid x_1 \in \mathbb{C}\}.$$

It is clear that $(0,0) \in U_1$. Furthermore, for $(x_1, 0), (y_1, 0) \in U_1$, we have

$$(x_1, 0) + (y_1, 0) = (x_1 + y_1, 0) \in U_1,$$

so U_1 is closed under addition. Similarly, for $a \in \mathbb{C}$ and $(x_1, 0) \in U_1$, we have

$$a(x_1, 0) = (ax_1, 0) \in U_1,$$

so U_1 is closed under scalar multiplication. Thus, U_1 is a subspace of V.

Additionally, let $U_2 = \{(0,0)\}$ and $U_3 = V$. It is clear that both U_2 and U_3 include the zero vector, are closed under addition, and are closed under scalar multiplication. Thus, U_1 and U_2 are both subspaces of V.

Finally, let
$$U_4 = \{(x_1, x_2) \in \mathbb{C}^2 \mid 2x_1 + x_2 = k\}.$$

For what values of k is U_4 a subspace? For $k \neq 0$, it follows that the zero vector is not in U_4, so it cannot be a subspace. For $k = 0$, the verification that U_4 is closed under addition and scalar multiplication is left as an exercise. Thus, U_4 is a subspace only if $k = 0$.

The next result gives an important property of subspaces.

Proposition 2.10
Suppose V is a vector space over $\mathbb{F}$ and $U \subset V$ is a subspace of V. Then, U is a vector space over $\mathbb{F}$ under the same addition and scalar multiplication of V.

Proof. To show that U is a vector space, we must show that it satisfies the six properties in Definition 2.1. The condition that $\vec{0} \in U$ ensures that the additive identity of V is in U. The closure of addition and scalar multiplication on U ensure that those two operations make sense over U.

Since U is closed under scalar multiplication, then $(-1)v = -v \in U$ for any $v \in U$. Thus, there exists an additive inverse for every vector in U. Finally, all other properties of a vector space are automatically satisfied in U because they hold in V. Thus, U is a vector space. $\square$

2.5 Sum of Subspaces

This section will introduce the notion of the sum of subspaces.

Definition 2.11 (sum of subspaces)
Suppose V is a vector space over $\mathbb{F}$ and $U_1, U_2 \subset V$ are subspaces of V. The **sum** of U_1 and U_2 is defined as

$$U_1 + U_2 = \{v \in V \mid v = u_1 + u_2 \text{ for some } u_1 \in U_1, u_2 \in U_2.\}$$

From this definition, it is clear that $U_1 + U_2$ is a subset of V.

Proposition 2.12
Suppose V is a vector space over $\mathbb{F}$ and $U_1, U_2 \subset V$ are subspaces of V. Then,

1. $U_1 + U_2$ is a subspace of V
2. $U_1 + U_2$ is the smallest subspace of V that contains both U_1 and U_2.

Proof. To prove (1), we must verify that $U_1 + U_2$ satisfies the three conditions in Definition 2.8. We know that

$\vec{0} \in U_1, U_2$, so

$$\vec{0} = \vec{0} + \vec{0} \in U_1 + U_2.$$

Now, suppose $v, w \in U_1 + U_2$. We can express $v = u_1 + u_2$ and $w = u_1' + u_2'$ for $u_1, u_1' \in U_1$ and $u_2, u_2' \in U_2$. Because U_1 is a subspace, it is closed under addition, so $u_1 + u_1' \in U_1$. Similarly, $u_2 + u_2' \in U_2$. Then,

$$v + w = (u_1 + u_2) + (u_2 + u_2') = (u_1 + u_1') + (u_2 + u_2') \in U_1 + U_2,$$

so $U_1 + U_2$ is closed under addition. Similarly, suppose $v = u_1 + u_2 \in U_1 + U_2$ and $a \in \mathbb{F}$. Because U_1 is a subspace, it is closed under scalar multiplication, so $au_1 \in U_1$. Similarly, $au_2 \in U_2$. Therefore,

$$av = a(u_1 + u_2) = au_1 + au_2 \in U_1 + U_2,$$

so $U_1 + U_2$ is closed under scalar multiplication. Thus, $U_1 + U_2$ is a subspace.

To prove (2), for all $u_1 \in U_1$, it follows that $u_1 + \vec{0} = u_1 \in U_1 + U_2$, so U_1 is contained in $U_1 + U_2$. Similarly, U_2 is contained in $U_1 + U_2$. Conversely, any subspace of V containing U_1 and U_2 must contain $U_1 + U_2$ because subspaces are closed under addition, so they must contain the sums of their elements. Therefore, $U_1 + U_2$ is the smallest subspace of V that contains both U_1 and U_2. □

Finally, we can extend this notion to the sum of an arbitrary number of subspaces $U_1 + \cdots + U_m$, for which Proposition 2.12 still holds.

3 Direct Sum, Span, and Linear Independence

3.1 Review

Last time, we introduced the notion of vectors and subspaces. We also introduced sums of subspaces.

3.2 Sum of Subspaces (continued)

Consider the following example of sum of subspaces.

Example 3.1
Suppose $V = \mathbb{C}^3$ and we define U_1 and U_2 as

$$U_1 = \{(x_1, x_2, x_3) \in \mathbb{C}^3 \mid x_1 + x_2 + x_3 = 0\}, \quad (1)$$
$$U_2 = \{(0, x, 0) \in \mathbb{C}^3 \mid x \in \mathbb{C}\}. \quad (2)$$

Then, U_1 and U_2 are subspaces (as an exercise, verify that this is the case). By the definition of sum of subspaces, any $v \in U_1 + U_2$ is of the form

$$v = (x_1, x_2, x_3) + (0, x, 0) = (x_1, x_2 + x, x_3)$$

such that $x_1 + x_2 + x_3 = 0$. To determine which vectors are in $U_1 + U_2$, let $v = (a, b, c) = (x_1, x_2 + x, x_3)$. Then, we have the system of equations

$$\begin{aligned} a &= x_1 \\ b &= x_2 + x \\ c &= x_3 \\ x_1 + x_2 + x_3 &= 0. \end{aligned}$$

We can easily solve this system to get

$$\begin{aligned} x_1 &= a \\ x_2 &= -a - c \\ x_3 &= c \\ x &= a + b + c. \end{aligned}$$

Because there exists a solution (x_1, x_2, x_3, x) for any $(a, b, c) \in \mathbb{C}^3$. it follows that any vector in $\mathbb{C}^3$ is also in $U_1 + U_2$. In particular,

$$(a, b, c) = (a, -a - c, c) + (0, a + b + c, 0) \in U_1 + U_2,$$

since $(a, -a - c, c) \in U_1$ and $(0, a + b + c, 0) \in U_2$. Therefore, $U_1 + U_2 = \mathbb{C}^3 = V$.

Student Question. *What if U_2 had a condition similar to U_1? For instance, define U_2' as*

$$U_2' = \{(x_1, x_2, x_3) \in \mathbb{C}^3 \mid 2x_1 + x_2 - 3x_3 = 0\}.$$

What is $U_1 + U_2'$?

Answer. *We use a similar process as the above example. Suppose $u_1 = (x_1, x_2, x_3) \in U_1$ and $u_2' = (y_1, y_2, y_3) \in U_2'$. Then, $v =\in U_1 + U_2$ if*

$$v = (a, b, c) = (x_1, x_2, x_3) + (y_1, y_2, y_3)$$

such that $x_1 + x_2 + x_3 = 0$ and $2y_1 + y_2 - 3y_2$. We can set up another system of equations including a, b, c to determine $U_1 + U_2$.

As a bit of an aside, consider the following example.

Example 3.2
Suppose $V = \mathbb{C}^3$ and define U_1 and U_2 the same as in Example 3.1. We wish to find $U_1 \cap U_2$. By the definitions of U_1 and U_2, we know that $v = (a, b, c) \in \mathbb{C}^3$ satisfies

- $a + b + c = 0$,
- $a = 0$, $c = 0$.

It is clear that the only solution is $(a, b, c) = (0, 0, 0)$, so $U_1 \cap U_2 = \{\vec{0}\}$. Thus, $U_1 \cap U_2$ is a subspace in this example.

In fact, any the intersection of any two subspaces of V is itself a subspace of V. The proof is left as an exercise.

3.3 Direct Sum

In this section, we will introduce the notion of direct sums.

Definition 3.3 (direct sum)
Suppose $U_1, \ldots, U_m$ are subspaces of V. We call $U_1 + \cdots + U_m$ a **direct sum** if every vector in $U_1 + \cdots + U_m$ can be written as $u_1 + \cdots + u_m$ in only one way, where each $u_i \in U_i$.

In other words, if $U_1 + \cdots + U_m$ is a direct sum and

$$v = u_1 + \cdots + u_m = w_1 + \cdots + w_m$$

for $u_i, w_i \in U_i$, then each $u_i = w_i$.

In terms of notation, if $U_1 + \cdots + U_m$ is a direct sum, then $U_1 \oplus \cdots \oplus U_m$ denotes $U_1 + \cdots + U_m$, with the $\oplus$ notation indicating that it is a direct sum. Additionally, we can use Σ notation to denote the sum of subspaces:

$$\sum_{i=1}^{m} U_i = U_1 + \cdots + U_m.$$

To denote a direct sum, we can use

$$\bigoplus_{i=1}^{m} U_i = U_1 \oplus \cdots \oplus U_m.$$

Now, consider the following example.

Example 3.4
Suppose $V = \mathbb{C}^3$ and define U_1 and U_2 the same as in Example 3.1. Is $U_1 + U_2$ a direct sum?

We wish to determine for any $v = (a, b, c) \in U_1 + U_2 = \mathbb{C}^3$ whether or not we can write $v = u_1 + u_2$ in more than one way, where $u_1 \in U_1$ and $u_2 \in U_2$. In Example 3.1, we found that the only way to express $v = u_1 + u_2$ is

$$v = (a, b, c) = (a, -a - c, c) + (0, a + b + c, 0).$$

Thus, $U_1 + U_2$ is a direct sum.

Student Question. *After Example 3.1, we defined*

$$U_2' = \{(x_1, x_2, x_3) \in \mathbb{C}^3 \mid 2x_1 + x_2 - 3x_3 = 0\}.$$

Is $U_1 + U_2'$ a direct sum?

Answer. *If we let $v = (a, b, c) \in U_1 + U_2'$, $u_1 = (x_1, x_2, x_3) \in U_1$, $u_2 = (y_1, y_2, y_3) \in U_2$, and set up a system of equations to describe $v = u_1 + u_2$, we would get a system of 5 homogeneous equations in 6 variables, which has infinitely many solutions. Thus, there is more than one way to express $v = u_1 + u_2$, so $U_1 + U_2'$ is not a direct sum. The details of this argument are left to the reader.*

The next result gives another condition to tell if a sum of subspaces is a direct sum.

Lemma 3.5
Suppose U_1, U_2 are subspaces of V. Then, $U_1 + U_2$ is a direct sum if and only if $U_1 \cap U_2 = \{\vec{0}\}$.

This is the first "if and only if"[3] statement in this course. To prove an "if and only if" statement, we must show that the statement holds in both directions.

Proof. First, we wish to prove that if $U_1 + U_2$ is a direct sum, then $U_1 \cap U_2 = \{\vec{0}\}$.[4] Suppose $v \in U_1 \cap U_2$. We know that

$$v = v + \vec{0}$$

where $v \in U_1$ and $\vec{0} \in U_2$. Similarly, we know

$$v = \vec{0} + v$$

where $\vec{0} \in U_1$ and $v \in U_2$. Since $U_1 + U_2$ is a direct sum, these two expressions must be the same, so $v = \vec{0}$. Thus, $U_1 \cap U_2 = \{\vec{0}\}$.

Now, we wish to prove that if $U_1 \cap U_2 = \{\vec{0}\}$, then $U_1 + U_2$ is a direct sum. Let $v \in U_1 \cup U_2$. Suppose we write v in two ways

$$v = u_1 + u_2 = w_1 + w_2$$

where $u_1, w_2 \in U_1$ and $u_2, w_2 \in U_2$. It remains to show that $u_1 = w_1$ and $u_2 = w_2$. We can rearrange the equation $u_1 + u_2 = w_1 + w_2$ to get

$$u_1 - w_1 = w_2 - u_2.$$

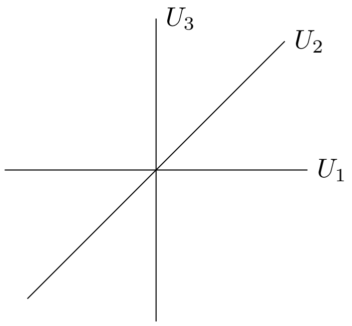
Since U_1 is a subspace, it is closed under addition, so $u_1 - w_1 \in U_1$. Similarly, $w_2 - u_2 \in U_2$. It follows that $u_1 - w_1 = w_2 - u_2 \in U_1 \cap U_2$, so $u_1 - w_1 = w_2 - u_2 = \vec{0}$. Therefore, $u_1 = w_1$ and $u_2 = w_2$, as desired. □

Example 3.6
Let us apply Lemma 3.5 to Example 3.4. By Example 3.2, we know that $U_1 \cap U_2 = \{\vec{0}\}$. Therefore, $U_1 + U_2$ is a direct sum.

A natural question is whether or not this result can be generalized for more than two subspaces; that is, if U_1, U_2, U_3 are subspaces of V, then does the condition

$$U_1 \cap U_2 = U_2 \cap U_3 = U_1 \cap U_3 = \{\vec{0}\}$$

imply that $U_1 + U_2 + U_3$ is a direct sum? To answer this question, consider the following subspaces of $\mathbb{R}^2$:



More precisely, we can define

$$U_1 = \{(x, 0) \mid x \in \mathbb{R}\},$$
$$U_2 = \{(0, y) \mid x \in \mathbb{R}\},$$
$$U_3 = \{(x, y) \mid x \in \mathbb{R}\}.$$

It is clear that the intersection of any two of these subspaces is the $(0, 0)$. However, note that

$$(1, 1) = (1, 0) + (0, 0) + (0, 1)$$

[3]The term "if and only if" also can be denoted by "i□" or with the double headed arrow " $\iff$."

[4]This is known as the *forward direction* of the proof, denoted by $\Rightarrow$. The converse of this statement is the *backwards direction*, denoted by $\Leftarrow$.

where $(1,0) \in U_1$, $(0,0) \in U_2$, and $(0,1) \in U_3$ and

$$(1,1) = (0,0) + (1,1) + (0,0)$$

where $(0,0) \in U_1$, $(1,1) \in U_2$, and $(0,0) \in U_3$. Thus, $U_1 + U_2 + U_3$ is not a direct sum.

Student Question. *We have just shown that the condition $U_1 \cap U_2 = U_2 \cap U_3 = U_1 \cap U_3 = \{\vec{0}\}$ is not su cient to prove that $U_1 + U_2 + U_3$ is a direct sum. However, is the condition $U_1 \cap U_2 = U_2 \cap U_3 = U_1 \cap U_3 = \{\vec{0}\}$* necessary *to prove that $U_1 + U_2 + U_3$ is a direct sum?*

Answer. *Yes! Without loss of generality, assume $U_1 \cap U_2 \neq \{\vec{0}\}$. By Lemma 3.5, $U_1 + U_2$ is not a direct sum, so $U_1 + U_2 + U_3$ is also not a direct sum.*

Student Question. *Is $U_1 + U_2 + U_3$ not a direct sum in this example because there are three subspaces but $V = \mathbb{R}^2$ is only two-dimensional?*

Answer. *We will introduce the notion of dimension in the next lecture, which will illuminate more properties about direct sums. In general, we will be able to show that three lines in $\mathbb{R}^2$ will never be a direct sum because the dimensions of the subspaces (lines) add up to more than the dimension of $\mathbb{R}^2$.*

We will now state the following result without proof.

Lemma 3.7
Suppose $U_1, \ldots, U_m$ are subspaces of V. Then, $U_1 + \cdots + U_m$ is a direct sum if and only if $\vec{0}$ can be written as $u_1 + \cdots + u_m$ ($u_i \in U_i$) in only one way, which is by taking all $u_i = \vec{0}$.

By Definition 3.3, if $U_1 + \cdots + U_m$ is a direct sum, then any $v \in U_1 + \cdots + U_m$ can be written uniquely as $u_1 + \cdots + u_m$ where $u_i \in U_i$. The above results tells us that if we can verify that if only the zero vector can be written uniquely in this way, then all vectors in $U_1 + \cdots + U_m$ can be written uniquely in this way.

3.4 Span

Now, we will shift our discussion to linear combinations and span.

Definition 3.8 (linear combination and span)
Suppose $v_1, \ldots, v_m$ are vectors in V. Then,

- a **linear combination** of $v_1, \ldots, v_m$ is a vector in V of the form

$$a_1 v_1 + \cdots a_m v_m$$

  for $a_1, \ldots, a_m \in \mathbb{F}$,
- the **span** of $v_1, \ldots, v_m$, denoted $\text{span}(v_1, \ldots, v_m)$ is the set of all linear combinations of $v_1, \ldots, v_m$.

Since all linear combinations of $v_1, \ldots, v_m$ are vectors in V, it follows that $\text{span}(v_1, \ldots, v_m)$ is a subset of V.

Example 3.9
Suppose $V = \mathbb{C}^3$ and

$$v_1 = (1, -1, 0), \quad v_2 = (0, 1, -1), \quad v_3 = (-1, 0, 1).$$

Then, any $v \in \text{span}(v_1, v_2, v_3)$ must be of the form

$$\begin{aligned} v &= a_1(1, -1, 0) + a_2(0, 1, -1) + a_3(-1, 0, 1) \\ &= (a_1 - a_3, -a_1 + a_2, -a_2 + a_3) \end{aligned}$$

for $a_1, a_2, a_3 \in \mathbb{C}$. To determine $\text{span}(v_1, v_2, v_3)$, we need to find all such v that can be expressed in this form. In particular, we notice that

$$(a_1 - a_3) + (-a_1 + a_2) + (-a_2 + a_3) = 0,$$

so the coordinates of any $v \in \text{span}(v_1, v_2, v_3)$ must add to 0. In fact, it is that case that

$$\text{span}(v_1, v_2, v_3) = \{(x_1, x_2, x_3) \in \mathbb{C}^2 \mid x_1 + x_2 + x_3 = 0\}.$$

The verification of the above statement is left as an exercise.

Let use clarify the span of an empty list of vectors.

Definition 3.10 (span())
The span of the empty list is $\{\vec{0}\}$.

Consider the following simple examples of span.

Example 3.11
These bases cases of span are useful to know:

- $\text{span}(\vec{0}) = \{\vec{0}\}$,
- $\text{span}(v, v, \ldots, v) = \text{span}(v) = \{av \mid a \in \mathbb{F}\}$.

3.5 Linear Independence

We can make an analogy between the span of vectors $\text{span}(v_1, \ldots, v_m)$ and the sum of subspaces $U_1 + \cdots + U_m$. In particular, we have the notion of direct sum $U_1 \oplus \cdots \oplus U_m$. Now, we will develop an analog of direct sum for vectors, with the notion of linear independence.

Definition 3.12 (linearly independent)
Suppose $v_1, \ldots, v_m \in V$. The list of vectors $v_1, \ldots, v_m$ is **linearly independent** if the equation

$$a_1 v_1 + \cdots + a_m v_m = \vec{0}$$

has the unique solution $a_1 = \cdots = a_m = 0$.

Note the similarity between the above definition to Lemma 3.7; both have a condition requiring a sum to be $\vec{0}$ where the only solution is when all variables are equal to 0.

Student Question. *I thought that the definition of linear independence is when vectors are not scalar multiples of each other. Is this correct?*

Answer. *This definition works to see if two vectors are linearly independent, but does not work for lists of more than two vectors.*

Consider the follow examples of linear independence.

Example 3.13
Suppose $V = \mathbb{C}^3$ and
$$v_1 = (1, -1, 0), \quad v_2 = (0, 1, -1), \quad v_3 = (-1, 0, 1).$$
Is v_1, v_2, v_3 linearly independent?

By observation, we have
$$v_1 + v_2 + v_3 = \vec{0}.$$
Therefore, v_1, v_2, v_3 is not linearly independent, also known as ***linearly dependent***.

Example 3.14
Suppose $v_1 = \vec{0}$ and $v_2 \neq \vec{0}$. Is v_1, v_2 linearly independent?

Note that
$$v_1 + 0v_2 = \vec{0},$$
which is different from $0v_1 + 0v_2 = 0$. Thus, v_1, v_2 is linearly dependent. In general, any list of vectors that contains $\vec{0}$ is linearly dependent.

Example 3.15
Suppose $V = \mathbb{C}^3$ and
$$v_1 = (1, 2, 3), \quad v_2 = (2, 3, 4), \quad v_3 = (3, 4, 5).$$
Is v_1, v_2, v_3 linearly independent?

It is easy to see by observation that
$$v_2 - v_1 = v_3 - v_2 = (1, 1, 1).$$
This implies that
$$2v_2 - v_1 - v_3 = \vec{0},$$
so v_1, v_2, v_3 is linearly dependent.

4 Basis and Dimension

4.1 Review

Last time, we introduced the definition

$$\text{span}(v_1, \dots, v_m) = \{\text{linear combinations of } v_1, \dots, v_m\}$$

where a linear combination of $v_1, \dots, v_m$ is a vector of the form $a_1v_1 + \cdots + a_mv_m$ for $a_i \in \mathbb{F}$. Note that $\text{span}(v_1, \dots, v_m)$ is a subspace of the original vector space V.

Furthermore, we defined a list $v_1, \dots, v_m$ to be linearly independent if $a_1v_1 + \cdots + a_mv_m = \vec{0}$ implies $a_1 = \cdots = a_m = 0$.

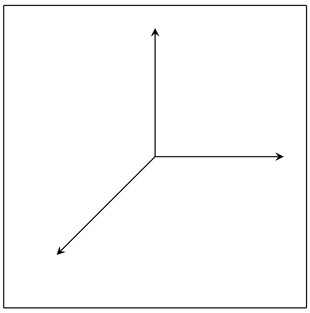
4.2 Span and Linear Independence (continued)

Suppose we choose enough vectors $v_1, \dots, v_m$ such that $\text{span}(v_1, \dots, v_m)$ is the entire vector space V.

Definition 4.1 (spans, spanning list)
If $\text{span}(v_1, \dots, v_m) = V$, we say that $v_1, \dots, v_m$ **spans** V or $v_1, \dots, v_m$ is a **spanning list** of V.

Intuitively, if we want to span a three-dimensional space, we need a list of at least 3 vectors to span the entire space. Soon, we will introduce the notion of dimension, and we will see that an n-dimensional vector space needs a list of at least n vectors to span the entire vector space.

Furthermore, we will see that a list of too many vectors cannot be linearly independent. In particular, a list of more than n vectors in an n-dimensional vector space cannot be linearly independent. For instance, the following three vectors in $\mathbb{R}^2$



$\mathbb{R}^2$

is linearly dependent.

4.3 Bases

Before we introduce bases, we will introduce a key definition in linear algebra.

Definition 4.2 (finite-dimensional vector space)
A vector space V is **finite-dimensional** if there exists a finite list of vectors $v_1, \dots, v_m$ that span V.

Note that this is just a qualitative definition, in the way that it doesn't actually tell you what the dimension of V is. For most of this course, we will assume that V is finite-dimensional.

Now, we will introduce another important definition.

Definition 4.3 (basis)
A **basis** of V is a list of vectors $v_1, \dots, v_m$ that both spans V and is linearly independent.

The spanning requirement makes sure the list of vectors is not too small, while the linearly independent requirement makes sure that the list is not too big. Eventually, we will show that the length of any basis of V is equal to the dimension of V.

Thus, the cardinality of a basis[5] is an ***invariant*** of V. In other words, if $v_1, \dots, v_m$ and $w_1, \dots, w_n$ are both bases of V, then $m = n$.

Consider the following example of a basis, known as the standard basis of $\mathbb{R}^n$.

Example 4.4
Recall $\mathbb{R}^n = \{(x_1, \dots, x_n) \mid x_i \in \mathbb{R}\}$. Let

$$\begin{aligned} e_1 &= (1, 0, \dots, 0), \\ e_2 &= (0, 1, 0, \dots, 0), \\ &\vdots \\ e_n &= (0, \dots, 0, 1). \end{aligned}$$

To check if $e_1, \dots, e_n$ is a basis, we need to check:

- $e_1, \dots, e_n$ span $\mathbb{R}^n$,
- $e_1, \dots, e_n$ are linearly independent.

To show that $e_1, \dots, e_n$ spans, we wish to find $a_1, \dots, a_n \in \mathbb{F}$ such that

$$(x_1, \dots, x_n) = a_1 e_1 + \cdots + a_n e_n$$

for any $(x_1, \dots, x_n) \in \mathbb{R}^n$. It is clear that this can be done by setting $a_i = x_i$. To prove linear independence, note that $a_1 e_1 + \cdots + a_n e_n = (a_1, \dots, a_n)$. Thus, if

$$a_1 e_1 + \cdots + a_n e_n = (a_1, \dots, a_n) = \vec{0},$$

then $a_1 = \cdots = a_n = 0$, so $e_1, \dots, e_n$ is linearly independent.

Therefore, $e_1, \dots, e_n$ is a basis of $\mathbb{R}^n$.

Consider a more abstract example of a basis.

Example 4.5
Suppose v_1, v_2 is a basis of V. We wish to show that $v_2, v_1 - v_2$ is also a basis of V.

First, we will show that $v_2, v_1 - v_2$ spans V. Since we know that v_1, v_2 spans V, then showing that v_1 and v_2 are linear combinations of $v_2, v1 - v_2$ is sufficient to prove that $v_2, v_1 - v_2$ spans V. Because $v_1 = v_2 + (v_1 - v_2)$ and $v_2 = v_2$, it follows that $v_2, v_1 - v_2$ spans V.

The verification that $v_2, v_1 - v_2$ is linearly independent is left as an exercise.

The above example shows us that if we are given a basis, we can make modifications to it (e.g. combinations, permutations) to make a new basis.

Student Question. *Shouldn't there be another condition in Definition 4.3, which is that the cardinality of any basis of V is invariant?*

Answer. *The two conditions in Definition 4.3 actually are enough to imply that the cardinality of any basis of V is invariant, which is a result we will prove later in this lecture.*

Now, we will introduce a technique that is very useful to proving the above statement.

[5] *Cardinality* denotes the number of elements. Thus, the cardinality of a basis is the number of vectors in the basis.

Algorithm 4.6 (Vector Deletion Algorithm)
The Vector Deletion Algorithm takes in a list of vectors $v_1, \dots, v_m$ and outputs a sublist $v_{i_1}, \dots, v_{i_l}$ ($1 \leq i_1 < \cdots < i_l \leq m$) such that

- $\text{span}(v_{i_1}, \dots, v_{i_l}) = \text{span}(v_1, \dots, v_m)$,
- $v_{i_1}, \dots, v_{i_l}$ is linearly independent.

The algorithm consists of scanning each vector exactly once:

- **Step 1**: If $v_1 = 0$, delete v_1, Otherwise, keep v_1.
- **Step j**: If $v_j \in \text{span}(v_1, \dots, v_{j-1})$, delete v_j. Otherwise, keep v_j.

Note that the condition $v_j \in \text{span}(v_1, \dots, v_{j-1})$ can be simplified to

$$v_j \in \text{span}(\text{current sublist of } v_1, \dots, v_{j-1}),$$

where the current sublist of $v_1, \dots, v_{j-1}$ represents all vectors remaining in the list after Step $j-1$ of the Vector Deletion Algorithm.

Student Question. *As an example, what if v_2 is in $\text{span}(v_5, v_6)$?*

Answer. *We don't care! At each step, we only look at the span of the previous vectors. In a later step, the algorithm would remove v_5 or v_6 (or both) from the list.*

Student Question. *What if we perform the Vector Deletion Algorithm on $\vec{0}, \dots, \vec{0}$?*

Answer. *We would delete v_1 since $v_1 = \vec{0}$. Additionally, $\vec{0}$ is in the span of any list of vectors (even the empty list), so we remove all vectors. Thus, the output would be the empty list.*

Consider the following example of the Vector Deletion Algorithm.

Example 4.7
Suppose $V = \mathbb{R}^2$. We wish to perform the Vector Deletion Algorithm on

$$\vec{0}, e_1, e_1 + e_2, e_1 - e_2$$

where e_1, e_2 is the standard basis of $\mathbb{R}^2$.

The algorithm deletes $v_1 = \vec{0}$, keeps $v_2 = e_1$ and $v_3 = e_1 + e_2$, and deletes $v_4 = e_1 - e_2 = 2e_1 - (e_1 + e_2) = 2v_2 - v_3$. Thus, the output is $e_1, e_1 + e_2$.

Now, we will use the Vector Deletion Algorithm to prove some results.

Proposition 4.8
Suppose $v_1, \dots, v_m$ span V. Then, there exists a sublist of $v_1, \dots, v_m$ that is a basis of V.

Proof. Apply VDA to $v_1, \dots, v_m$. The output will be a sublist $v_{i_1}, \dots, v_{i_l}$ such that $\text{span}(v_{i_1}, \dots, v_{i_l}) = \text{span}(v_1, \dots, v_m) = V$ and $v_{i_1}, \dots, v_{i_l}$ is linearly independent. Therefore, $v_{i_1}, \dots, v_{i_l}$ is a basis of V. □

Note that while we claimed that the output of the Vector Deletion Algorithm is always linearly independent, we haven't rigorously proved it yet. Thus, we will prove it now.

Lemma 4.9
The output of the Vector Deletion Algorithm is linearly independent.

Proof. We will prove this by contradiction. Suppose the output $v_{i_1}, \dots, v_{i_l}$ is linearly dependent. It follows that $a_1 v_{i_1} + \cdots + a_l v_{i_l} = \vec{0}$ for scalars $a_1, \dots, a_l \in \mathbb{F}$ that are not all zero. Let a_r be the last nonzero coefficient in $a_1, \dots, a_l$. Since the later coefficients are all zero, it follows that

$$a_1 v_{i_1} + \cdots + a_r v_{i_r} = \vec{0}.$$

We can solve for v_{i_r} to get

$$v_{i_r} = -\frac{1}{a_r}(a_1 v_{i_1} + \cdots + a_{r-1} v_{i_{r-1}}).$$

From the above equation, it is clear that $v_{i_r} \in \text{span}(v_{i_1}, \ldots, v_{i_{r-1}})$. However, this implies that v_{i_r} should have been deleted by VDA, which is a contradiction. Therefore, $v_{i_1}, \ldots, v_{i_l}$ is linearly independent. □

Another property of the Vector Deletion Algorithm is that if $v_1, \ldots, v_r$ are linearly independent, then the output of VDA also starts with $v_1, \ldots, v_r$. In other words, no vector among $v_1, \ldots, v_r$ is deleted by the VDA. This can be seen by the fact that if $v_1, \ldots, v_j$ is linearly independent, then $v_j \notin \text{span}(v_1, \ldots, v_{j-1})$.

We will use the above property to prove the following result.

Proposition 4.10
Suppose $v_1, \ldots, v_m$ is linearly independent. Then, there exists $v_{m+1}, \ldots, v_n$ such that $v_1, \ldots, v_m, v_{m+1}, \ldots, v_n$ is a basis of V.

Proof. Since V is finite-dimensional,[6] there exists vectors $w_1, \ldots, w_k$ that span V. Then, the list

$$v_1, \ldots, v_m, w_1, \ldots, w_k$$

spans V. Therefore, applying VDA to this list gives an output that spans V and is linearly independent, so the output is a basis of V. Since $v_1, \ldots, v_m$ is linearly independent, it follows that the output of VDA must also contain $v_1, \ldots, v_m$. Finally, renaming the remaining vectors in the output to $v_{m+1}, \ldots, v_n$ gives the desired result. □

Note the similarities between Proposition 4.8 and Proposition 4.10. In particular, Proposition 4.8 states that if we have a spanning list (possibly too many vectors for a basis), then we can delete some vectors to form a basis of V. On the other hand, Proposition 4.10 stats that if we have a linearly independent list of vectors (possibly too few vectors), then we can add some vectors to form a basis of V.

4.4 Dimension

Now, we are ready to prove that the number of vectors in a basis is invariant.

Theorem 4.11
Suppose $v_1, \ldots, v_m$ and $w_1, \ldots, w_n$ are both bases of V. Then, $m = n$.

Proof. Let $B_0 = (v_1, \ldots, v_m)$. Add w_1 to this list to form $B_1^+ = (w_1, v_1, \ldots, v_m)$. Then, B_1^+ spans V, but B_1^+ is not linearly independent because $v_1, \ldots, v_m$ already spans V, so w_1 can be expressed as a linear combination of $v_1, \ldots, v_m$. Apply VDA to B_1^+ to obtain output B_1. By Proposition 4.8, B_1 is a basis. Also, since $w_1 \neq 0$, B_1 must contain w_1. Finally, since B_1^+ is not linearly independent, VDA must delete at least one vector, so $|B_1| \leq m$.[7]

Now, add w_2 to this list to form $B_2^+ = (w_2, B_1)$. By similar logic, B_2^+ spans V and is linearly dependent. Thus, applying VDA to B_2^+ gives output B_2 that is a basis and starts with w_2, w_1, because all the w_i's are linearly independent. Once again, VDA must delete at least one vector, so $|B_2| \leq |B_1| \leq m$.

Repeating this process n times, we find that B_n is a basis starting with $w_n, \ldots, w_1$. However, $w_n, \ldots, w_1$ is already a basis, so B_n contains no other vectors. Furthermore, $|B_n| \leq m$, so $n \leq m$.

Finally, reversing the roles of $v_1, \ldots, v_m$ and $w_1, \ldots, w_n$ (by letting $B_0 = (w_n, \ldots, w_n)$ and adding v_i's) shows that $m \leq n$. Therefore, $m = n$. □

Now, we have a well-defined definition of dimension.

[6] Unless stated otherwise, we will assume V is finite-dimensional for all results in this course.
[7] The notation $|B_1|$ represents the number of elements in B_1.

Definition 4.12 (dimension)
The **dimension** of V, denoted $\dim V$, is n if there exists a basis of V consisting of n vectors.

We can also use the notation $\dim_{\mathbb{F}} V$ to emphasize the dimension over field $\mathbb{F}$, where $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

4.5 Polynomials

While we ran out of time in the lecture, the following concepts on polynomials will show up often in the course.

Definition 4.13 (polynomial)
A function $p : \mathbb{F} \to \mathbb{F}$ is called a **polynomial** if there exist scalars $a_0, \ldots, a_m \in \mathbb{F}$ such that

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

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

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

Definition 4.14 (degree of a polynomial)
A polynomial $p \in \mathcal{P}(\mathbb{F})$ has **degree** m, denoted $\deg m = p$, if there exist scalars $a_0, \ldots, a_m \in \mathbb{F}$ with $a_m \neq 0$ such that

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

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

The zero polynomial is defined to have degree $-\infty$

Definition 4.15 ($\mathcal{P}_m(\mathbb{F})$)
For nonnegative integer m, $\mathbb{P}_m(\mathbb{F})$ denotes the set of all polynomials with coefficients in $\mathbb{F}$ and degree at most m.

Example 4.16
$\mathcal{P}_m(\mathbb{F})$ is a finite-dimensional vector space for any nonnegative integer m.

5 Dimension (continued), Linear Maps, and Matrices

5.1 Review

Last time, we introduced the notion of a basis. In particular, a list of vectors $v_1, \ldots, v_n$ is a basis of V if

- $\operatorname{span}(v_1, \ldots, v_n) = V$,
- $v_1, \ldots, v_n$ is linearly independent.

Additionally, we showed that the number of vectors in a basis is an invariant of V. Thus, we defined the dimension of V as the number of vectors in any basis of V.

5.2 Dimension (continued)

Consider the following basic examples of dimension.

Example 5.1
Consider $\mathbb{F}^n$. We have $\dim \mathbb{F}^n = n$ with one basis being the standard basis of $\mathbb{F}^n$.

Recall the standard basis is the list of vectors $e_1, \ldots, e_n$, where e_i has 1 at the i^{th} place and 0's elsewhere.

Example 5.2
Recall

$$\mathcal{P}_m(\mathbb{F}) = \{a_0 + a_1 z + \cdots + a_m z^m \mid a_i \in \mathbb{F}\}.$$

In other words, $\mathcal{P}_m(\mathbb{F})$ is the set of all polynomials with degree at most m.

We claim that $1, z, z^2, \ldots, z^m$ is a basis of $\mathcal{P}_M(\mathbb{F})$. It is clear that $1, z, z^2, \ldots, z^m$ span by the definition of $\mathcal{P}_m(\mathbb{F})$. To prove linear independence, we need to show that $a_0 + a_1 z + \cdots + a_m z^m = 0$ implies $a_0 = a_1 = \cdots = a_m = 0$. It is a known fact in algebra that a polynomial of degree m can have at most m distinct roots, so it cannot vanish for all values of z. Therefore, $a_0 = a_1 = \cdots = a_m = 0$.

Thus, $1, z, z^2, \ldots, z^m$ forms a basis, so $\dim \mathcal{P}_m(\mathbb{F}) = m + 1$.

Now, consider the following result relating dimension to spanning lists and linear independence.

Proposition 5.3
Suppose $v_1, \ldots, v_n$ is a list of vectors in V. Then,

1. If $v_1, \ldots, v_n$ is linearly independent, then $n \leq \dim V$. Moreover, if $n = \dim V$, then $v_1, \ldots, v_n$ is a basis.
2. If $v_1, \ldots, v_n$ span V, then $n \geq \dim V$. Moreover, if $n = \dim V$, then $v_1, \ldots, v_n$ is a basis.

Proof. To prove (1), recall by Proposition 4.10 that $v_1, \ldots, v_n$ can be extended to a basis $v_1, \ldots, v_n, v_{n+1}, \ldots, v_m$ of V. It follows that $\dim V = m \geq n$. Furthermore, if $m = n$, then $v_1, \ldots, v_n$ is already a basis of V.

To prove (2), recall by Proposition 4.8 that there exists a sublist $v_{i_1}, \ldots, v_{i_r}$ that is a basis of V. It follows that $\dim V = r \leq n$. Furthermore, if $r = n$, then $v_1, \ldots, v_n$ is already a basis of V. □

This result is very useful because it asserts that we only need to verify one of the two conditions of a basis (linear independence and span) given that a list of vectors is the right length. The next example will demonstrate this.

Example 5.4
Consider the polynomials 1, z, $z(z-1)$, $z(z-1)(z-2), \dots, z(z-1)\cdots(z-(m-1))$. Is this a basis of $\mathcal{P}_m(\mathbb{F})$?

First, we can consider some simple cases.

- For $m=1$, we have the list $1, z$, which is a basis.
- For $m=2$, we have $1, z, z(z-1)$. To prove linear independence, suppose

$$a + bz + cz(z-1) = 0$$

for scalars $a, b, c \in \mathbb{F}$. Expanding, we get $cz^2 + (b-c)z + a = 0$, which has only solution $a = b = c = 0$. Therefore, $1, z, z(z-1)$ is linearly independent. Additionally, we can show that

$$a + bz + cz(z-1) = a_0 + a_1 z + a_2 z^2$$

has solution $(a, b, c) = (a_0, a_1 + a_2, a_2)$, so $1, z, z(z-1)$ spans V. Therefore, this list is a basis.

Now, we will return to the general case. We claim that 1, z, $z(z-1)$, $z(z-1)(z-2), \dots, z(z-1)\cdots(z-(m-1))$ is linearly independent. For the sake of contradiction, suppose

$$a_0 + a_1 z + a_2 z(z-1) + \cdots + a_m z(z-1)\cdots(z-(m-1)) = 0$$

for scalars $a_0, a_1, \dots, a_m \in \mathbb{F}$ that are not all zero. Let a_j be the last nonzero coefficient of $a_0, a_1, \dots, a_m$. It follows that

$$a_0 + a_1 z + \cdots + a_j z(z-1)\cdots(z-(j-1)) = 0.$$

By observation, we see that the coefficient of the z^j term is a_j. However, since $a_j \neq 0$ it follows that the left-hand side is nonzero, which is a contradiction. Therefore, 1, z, $z(z-1)$, $z(z-1)(z-2), \dots, z(z-1)\cdots(z-(m-1))$ is linearly independent.

Finally, since the number of vectors in 1, z, $z(z-1)$, $z(z-1)(z-2), \dots, z(z-1)\cdots(z-(m-1))$ is equal to $\dim \mathcal{P}_m(\mathbb{F})$, it follows that 1, z, $z(z-1)$, $z(z-1)(z-2), \dots, z(z-1)\cdots(z-(m-1))$ is a basis by Proposition 5.3.

Now, we will look at the dimension of a sum of subspaces.

Proposition 5.5
Suppose U_1, U_2 are subspaces of V. Then,

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

Proof. Suppose $u_1, \dots, u_n$ is a basis of $U_1 \cap U_2$. Since this list is linearly independent, we can extend this list to a basis $u_1, \dots, u_n, v_1, \dots, v_m$ of U_1. Similarly, we can extend this list to a basis $u_1, \dots, u_n, w_1, \dots, w_k$ of U_2.

Now, we claim that $u_1, \dots, u_n, v_1, \dots, v_m, w_1, \dots, w_k$ is a basis of $U_1 + U_2$. This will give us

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

which will complete the proof.

It is clear that $\operatorname{span}(u_1, \dots, u_n, v_1, \dots, v_m, w_1, \dots, w_k)$ contains both U_1 and U_2, and thus also contains $U_1 + U_2$. To prove linear independence, suppose

$$a_1 u_1 + \cdots + a_n u_n + b_1 v_1 + \cdots + b_m v_m + c_1 w_1 + \cdots + c_k w_k = 0$$

where all a_i's, b_i's, and c_i's are scalars. Rearranging, we get

$$c_1 w_1 + \cdots + c_k w_k = -a_1 u_1 - \cdots - a_n u_n - b_1 v_1 - \cdots - b_m v_m.$$

All the w_i's are in U_2, so $c_1w_1+\cdots+c_kw_k \in U_2$. Furthermore, all u_i's and v_i's are in U_1, so $c_1w_1+\cdots+c_kw_k \in U_1$ as well. It follows that $c_1w_1+\cdots+c_kw_k \in U_1 \cap U_2$. Since $u_1, \ldots, u_n$ is a basis of $U_1 \cap U_2$, we can write

$$c_1w_1 + \cdots + c_kw_k = d_1u_1 + \cdots + d_nu_n$$

for scalars $d_1, \ldots, d_n$. However, $u_1, \ldots, u_n, w_1, \ldots, w_k$ is linearly independent because it forms a basis, so all c_i's and d_i's must be equal to 0. In particular, all c_i's are equal to 0, so we have

$$a_1u_1 + \cdots + a_nu_n + b_1v_1 + \cdots + b_mv_m = 0.$$

However, $u_1, \ldots, u_n, v_1, \ldots, v_m$ is also linearly independent because it forms a basis, so all a_i's and b_i's are 0. Therefore, $u_1, \ldots, u_n, v_1, \ldots, v_m, w_1, \ldots, w_k$ is linearly independent and thus is a basis of U_1+U_2, as desired. □

This result is analogous to the inclusion-exclusion principle in discrete math, which states that $|X_1 \cup X_2| = |X_1| + |X_2| - |X_1 \cap X_2|$ for sets X_1, X_2. Furthermore, note that we cannot use Proposition 5.3 for the above proof because $\dim(U_1 + U_2)$ is unknown.

A special case of the above result is when $U_1 + U_2$ forms a direct sum, $U_1 \oplus U_2$. This implies that $U_1 \cap U_2 = \{\vec{0}\}$, which has dimension 0.[8] Thus, $\dim(U_1 \oplus U_2) = \dim U_1 + \dim U_2$.

Student Question. *Does Proposition 5.5 generalize to the sum of n vector spaces, $U_1 + \cdots + U_n$?*

Answer. *Yes, the formula actually takes the same form as the inclusion-exclusion principle. You can read about the inclusion-exclusion principle more here: https://en.wikipedia.org/wiki/Inclusion%E2%80%93exclusion_principle.*

5.3 Linear Maps

So far, we have focused on vector spaces. Now, we will discuss how we can relate two different vector spaces, using linear maps.

Definition 5.6 (linear map)
Suppose V, W are vector spaces over $\mathbb{F}$. A map $T : V \to W$ is called a **linear map** if it satisfies:

- additivity:
$$T(v_1 + v_2) = T(v_1) + T(v_2) \text{ for all } v_1, v_2 \in V;$$
- homogeneity:
$$T(cv) = c \cdot T(v) \text{ for all } c \in \mathbb{F} \text{ and all } v \in V.$$

Note that the equation $T(v_1 + v_2) = T(v_1) + T(v_2)$ uses two different additions; the left-hand side uses abstract addition in V while the right-hand side uses abstract addition in W. Similarly, the left-hand side of $T(cv) = c \cdot T(v)$ uses abstract scalar multiplication in V while the right-hand side uses abstract scalar multiplication in W.

Example 5.7
Suppose $T : \mathbb{R}^2 \to \mathbb{R}^3$ is defined by

$$T(x, y) = (3x - y, x + y, 5x - y).$$

Then, T is a linear map, the verification of which is left as an exercise. On the other hand, suppose

$$T(x, y) = (3x - y - 1, x + y, 5x - y).$$

While $3x - y - 1$ may be considered a linear function in calculus, T is actually not a linear map.

To show that $T(x, y) = (3x - y - 1, x + y, 5x - y)$ is not a linear map, consider the following result.

Lemma 5.8
Suppose $T : V \to W$ is a linear map. Then, $T(\vec{0}) = \vec{0}$.

[8] The zero vector space $\{\vec{0}\}$ has the empty set as a basis, which is why $\dim\{\vec{0}\} = 0$.

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

$$T(\vec{0}) = T(0v) = 0 \cdot T(v) = \vec{0},$$

as desired. □

We see that $T(0,0) = (-1,0,0)$. Thus, by the above result, T is not a linear map.

Now, we will investigate more examples of linear maps.

Example 5.9
Consider the following simple examples of linear maps:

1. The zero map $0 : V \to W$ defined by $0(v) = \vec{0}$ for all $v \in V$ is a linear map.
2. The identity map $\mathrm{id}_V : V \to V$ defined by $\mathrm{id}_V(v) = v$ for all $v \in V$ is a linear map.
3. For $c \in \mathbb{F}$, the map $T : V \to V$ defined by $T(v) = cv$ for all $v \in V$ is a linear map.
4. The map $T : \mathcal{P}_m(\mathbb{F}) \to \mathcal{P}_{m+1}(\mathbb{F})$ defined by $T(p(z)) = zp(z)$. We can verify that

$$T(p(z) + q(z)) = z(p(z) + q(z)) = zp(z) + zq(z) = T(p(z)) + T(q(z))$$

   and

$$T(c \cdot p(z)) = z \cdot c \cdot p(z) = c(zp(z)) = c \cdot T(p(z)),$$

   so T is linear.
5. The map $T : \mathcal{P}_m(\mathbb{F}) \to \mathcal{P}_{m+1}(\mathbb{F})$ defined by $T(p(z)) = p'(z)$. We can verify that

$$T(p + q) = (p + q)' = p' + q' = T(p) + T(q)$$

   and

$$T(cp) = (cp)' = cp' = c \cdot T(p),$$

   so T is linear.
6. The map $T : \mathcal{P}_m(\mathbb{F}) \to \mathbb{F}$ defined by $T(p(z)) = p(1)$. We can verify that

$$T(p + q) = (p + q)(1) = p(1) + q(1) = T(p) + T(q)$$

   and

$$T(cp) = (cp)(1) = c \cdot p(1) = c \cdot T(p),$$

   so T is linear.

5.4 Matrices

Consider the linear map from Example 5.7:

$$T(x, y) = (3x - y, x + y, 5x - y).$$

We wish to represent this linear map using a matrix, which we do by taking the coefficients on the right-hand side and putting them into each of the rows:

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

Thus, we represent a linear map $T : \mathbb{R}^2 \to \mathbb{R}^3$ with a 3×2 matrix.

Now, let's construct this matrix in a different way. Let us calculate where T maps each of the basis vectors e_1, e_2 of $\mathbb{R}^2$:

$$\begin{aligned} Te_1 &= T(1,0) = (3,1,5) \\ Te_2 &= T(0,1) = (-1,1,-1). \end{aligned}$$

Then, putting these vectors into the columns gives the same 3×2 matrix

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

In general, the latter method allows us to find the matrix representation of any linear map with abstract vector spaces $T : V \to W$. In particular, the matrix will be of size $\dim W \times \dim V$.

First, we will consider a simpler case where V, W are not completely generalized.

Example 5.10
Suppose $T : \mathbb{F}^n \to \mathbb{F}^m$ is a linear map. The matrix representation of T is

$$\begin{bmatrix} T(e_1) & \cdots & T(e_n) \end{bmatrix},$$

where $e_1, \ldots, e_n$ is the standard basis of $\mathbb{F}^n$ and $T(e_1), \ldots, T(e_n)$ are written as column vectors of length m. Thus, this matrix has size $m\times n$.

In the general case where $T : V \to W$ is a linear map, choose a basis $v_1, \ldots, v_n$ of V and a basis $w_1, \ldots, w_m$ of W. Then, $T(v_i) \in W$, so it can be expressed as a linear combination of w_i's:

$$\begin{aligned} T(v_1) &= a_{11}w_1 + a_{21}w_2 + \cdots + a_{m1}w_m \\ T(v_2) &= a_{12}w_1 + a_{22}w_2 + \cdots + a_{m2}w_m \\ &\vdots \\ T(v_2) &= a_{1n}w_1 + a_{2n}w_2 + \cdots + a_{mn}w_m. \end{aligned}$$

Then, a matrix representation of T, denoted $\mathcal{M}(T)$, is

$$\mathcal{M}(T) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

Thus, the general rule to express $T(v_i)$ as a linear combination of $w_1, \ldots, w_m$ and put the coordinates of $T(v_i)$ as the i^{th} column of $\mathcal{M}(T)$.

Example 5.11
Suppose $T : \mathcal{P}_3(\mathbb{F}) \to \mathcal{P}_2(\mathbb{F})$ defined by $T(p(z)) = p'(z)$. By Example 5.9, T is a linear map. We wish to find $\mathcal{M}(T)$.

We choose $1, z, z^2, z^3$ as a basis of $\mathcal{P}_3(\mathbb{F})$ and $1, z, z^2$ as a basis of $\mathcal{P}_3(\mathbb{F})$. Then, we calculate T applied to each of the basis vectors of $\mathcal{P}_3(\mathbb{F})$:

$$\begin{aligned} T(1) &= 0 = 0 \cdot 1 + 0z + 0z^2 \\ T(z) &= z' = 1 = 1 \cdot 1 + 0z + 0z^2 \\ T(z^2) &= (z^2)' = 2z = 0 \cdot 1 + 2z + 0z^2 \\ T(z^3) &= (z^3)' = 3z^2 = 0 \cdot 1 + 0z + 3z^2. \end{aligned}$$

Putting the coefficients of the $1, z, z^2$ terms into the columns of a matrix, we find that

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

with respect to the chosen bases. However, note that $\mathcal{M}(T)$ is dependent on choice of basis. For instance, suppose we choose $1, z, z(z-1)$ as a basis of $\mathcal{P}_2(\mathbb{F})$. Then,

$$\begin{aligned} T(1) &= 0 \cdot 1 + 0z + 0z^2 \\ T(z) &= 1 \cdot 1 + 0z + 0z^2 \\ T(z^2) &= 0 \cdot 1 + 2z + 0z^2 \\ T(z^3) &= 0 \cdot 1 + 3z + 3z^2. \end{aligned}$$

Thus, we find that the matrix representation of T with respect to this basis is

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

6 Matrices (continued) and Matrix Multiplication

6.1 Review

Last time, we introduced he notion of linear maps and related them with matrices. Suppose $T : V \to W$ is a linear map. Then, T satisfies

- $T(v_1 + v_2) = T(v_1) + T(v_2)$,
- $T(\lambda v) = \lambda T(v)$ for $\lambda \in \mathbb{F}$.

Also, an $m \times n$ matrix takes the form

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \vdots & \vdots & & \vdots \\ A_{m,1} & A_{m,2} & \cdots & A_{m,n} \end{pmatrix},$$

where m is the number of rows and n is the number of columns.[9]

We can turn a linear map into a matrix by the following steps:

1. Choose a basis $v_1, \ldots, v_n$ of V and a basis $w_1, \ldots, w_m$ of W.
2. Express $T(v_i)$ as a linear combination of w_j's.
3. Put these coefficients into the columns of a matrix:

$$\mathcal{M}(T) = \begin{bmatrix} T(e_1) & \cdots & T(e_n) \end{bmatrix}.$$

6.2 Matrices Under Different Bases

Recall that the matrix of a linear map $T : V \to W$ depends on the chosen bases of V and W. In this section, we will investigate matrices representing the same linear map under different bases.

In all of the following examples, suppose we have the matrix of a linear map $T : V \to W$ with respect to bases $v_1, \ldots, v_n$ of V and $w_1, \ldots, w_m$ of W.

Example 6.1
Suppose we wish to find the matrix of T using $w_2, w_1, w_3, \ldots, w_m$ as a basis of W. We have

$$\begin{aligned} T(v_1) &= c_1 v_1 + c_2 w_2 + \cdots c_n w_n \\ &= c_2 v_2 + c_1 w_1 + \cdots c_n w_n. \end{aligned}$$

Thus, we can express the first column of $\mathcal{M}(T)$ with respect to $w_2, w_1, w_3, \ldots, w_m$ as

$$\begin{pmatrix} c_2 & \cdots & \cdots \\ c_1 & & \\ c_3 & \cdots & \cdots \\ \vdots & & \\ c_n & \cdots & \cdots \end{pmatrix}.$$

Extending this logic to all $T(v_i)$, we see that

$$\mathcal{M}(T, (w_2, w_1, w_3, \ldots, w_m)) = \text{swap first two rows of } \mathcal{M}(T, (w_1, \ldots, w_m)).$$

[9] Note that a $1 \times n$ matrix is often called a row vector and a $n \times 1$ matrix is often called a column vector.

Example 6.2
Similarly, suppose we wish to find the matrix of T with respect to the basis $v_2, v_1, v_3, \ldots, v_n$ of V. By similar logic, we can show that

$$\mathcal{M}(T,(v_2,v_1,v_3,\ldots,v_n)) = \text{swap first two rows of } \mathcal{M}(T,(v_1,\ldots,v_n)).$$

Example 6.3
Suppose we wish to find the matrix of T using $w_1+w_2, w_2, w_3, \ldots, w_m$ as a basis of W. We have

$$\begin{aligned} T(v_1) &= c_1w_1 + c_2w_2 + \cdots + c_mw_m \\ &= c_1w_1 + (c_2-c_1)c_2 + \cdots + c_mw_m. \end{aligned}$$

It follows that the first column of $\mathcal{M}(T)$ with respect to $w_1+w_2, w_2, w_3, \ldots, w_m$ is

$$\begin{pmatrix} c_1 & \cdots & \cdots \\ c_2-c_1 & & \\ c_3 & \cdots & \cdots \\ \vdots & & \\ c_n & \cdots & \cdots \end{pmatrix}.$$

The same logic follows for all columns, so

$$\mathcal{M}(T,(w_1+w_2,w_2,w_3,\ldots,w_m)) = \text{subtract } 1^{\text{st}} \text{ column from } 2^{\text{nd}} \text{ column of } \mathcal{M}(T,(w_1,\ldots,w_m)).$$

There is no need to remember the results from any of these examples; instead, these examples are meant to give intuition on how to express matrices with respect different bases.

6.3 Matrices and Linear Maps

In terms of notation, we write $\mathbb{F}^{m,n}$ to denote the set of all $m\times n$ matrices with entries in $\mathbb{F}$. We also write $\mathcal{L}(V,W)$ to represent the set of all linear operators from V to W.

We have spent the last few sections showing how we can express any linear map $T \in \mathcal{L}(V,W)$ as a matrix $\mathcal{M}(T) \in \mathbb{F}^{m,n}$. If we fix bases (v_i) and (w_j) of V and W, respectively, this creates a map $\mathcal{L}(V,W) \to \mathbb{F}^{m,n}$.[10]

Guiding Question
Suppose $T \in \mathcal{L}(V,W)$ has matrix representation $\mathcal{M}(T) \in \mathbb{F}^{m,n}$. What are m and n in terms of the dimensions of V and W?

Answer. *The columns of $\mathcal{M}(T)$ is the coordinates of vectors in W. Thus, the number of rows should be $m = \dim W$ and the number of columns should be $n = \dim V$. This question should hopefully be review.*

Now, we will prove an important result relating $\mathcal{L}(V,W)$ and $\mathbb{F}^{m,n}$.

Proposition 6.4
The map that sends a linear operator $T \in \mathcal{L}(V,W)$ to its matrix representation $\mathcal{M}(T) \in \mathbb{F}^{m,n}$ is a bijection.

Proof. To prove that this map is a bijection, it is sufficient to construct an inverse map. Thus, for any matrix $A \in \mathbb{F}^{m,n}$, we wish to construct a map that sends A to a linear operator $T_A \in \mathcal{L}(V,W)$ such that $\mathcal{M}(T_A) = A$. Suppose $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$ are bases of V and W, respectively. Note that the condition $\mathcal{M}(T_A) = A$

[10] To emphasize the bases of V and W when writing the matrix of an operator, we write $\mathcal{M}(T,(v_i),(w_j))$ to denote the matrix of $T \in \mathcal{L}(V,W)$ with respect to bases (v_i) and (w_j) of V and W, respectively.

implies that

$$\begin{aligned}
T_A(v_1) &= A_{11}w_1 + \cdots A_{m1}w_m \\
T_A(v_2) &= A_{12}w_1 + \cdots A_{m2}w_m \\
&\vdots \\
T_A(v_n) &= A_{1n}w_1 + \cdots A_{mn}w_m.
\end{aligned}$$

The right-hand side of each equation is some arbitrary vector in W, which we can denote $u_1, \ldots, u_n$:

$$T_A(v_1) = u_1, \quad T_A(v_2) = u_2, \quad \ldots, \quad T_A(v_n) = u_n.$$

Now, we will show that T_A exists. Since $v_1, \ldots, v_n$ is a basis, we can express any $v \in V$ as $v = a_1v_1 + \cdots + a_nv_n$ for scalars $a_1, \ldots, a_n$. Since T_A is linear, we have

$$\begin{aligned}
T_A(v) &= T_A(a_1v_1 + \cdots + a_nv_n) \\
&= a_1T(v_1) + \cdots + a_nT(v_n) \\
&= a_1u_1 + \cdots + a_nu_n.
\end{aligned}$$

Since $a_1, \ldots, a_n$ are uniquely determined, it follows that T_A maps any $v \in V$ to exactly one vector in W, so T_A exists.[11]

Now, we will show that the map that sends A to T_A is an inverse of T. To do this, we need to show:

- $T_A = T$ for $T \mapsto \mathcal{M}(T) = A \mapsto T_A$,
- $\mathcal{M}(T_A) = A$ for $A \mapsto T_A \mapsto \mathcal{M}(T_A)$

We will only prove the first point above; the second follows similarly and is left as an exercise for the reader. We know that

$$T_A(v_i) = A_{1i}w_1 + \cdots A_{mi}w_m.$$

However, we also know that the i^{th} column of A is the coordinates of $T(v_i)$ with respect to $w_1, \ldots, w_m$. In other words,

$$T(v_i) = A_{1i}w_1 + \cdots A_{mi}w_m.$$

Since $v_1, \ldots, v_n$ is a basis, it follows that $T_A = T$, as desired. □

The above result shows that the abstract set of linear maps $\mathcal{L}(V, W)$ can be concretely identified by $m \times n$ matrices.

Note that $\mathbb{F}^{m,n}$ is simply a rectangular grid of mn numbers. Thus, we can view $\mathbb{F}^{m,n}$ as a vector space with dimension mn, with addition defined as element-wise addition and scalar multiplication defined as element-wise scalar multiplication. To put it more rigorously, addition is defined as $(A + B)_{ij} = A_{ij} + B_{ij}$ and scalar multiplication is defined as $(cA)_{ij} = cA_{ij}$.

Now, since $\mathcal{L}(V, W)$ and $\mathbb{F}^{m,n}$ are bijective by Proposition 6.4, we can also view $\mathcal{L}(V, W)$ as a vector space, so we define addition and scalar multiplication in $\mathcal{L}(V, W)$. Suppose $S, T \in \mathcal{L}(V, W)$, so S, T are linear maps from V to W. We can define addition as

$$(S + T)(v) = S(v) + T(v).$$

Note that the + symbol on left-hand side represents abstract addition in $\mathcal{L}(V, W)$, while the + symbol on the right-hand side represents abstract addition in W. Similarly, for $c \in \mathbb{F}$, we define scalar multiplication as

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

Once again, note that the left-hand side uses scalar multiplication in $\mathcal{L}(V, W)$, while the right-hand side uses scalar multiplication in W. The verification that $\mathcal{L}(V, W)$ is a vector space with respect to these definitions of addition and scalar multiplication is left as an exercise to the reader.

This gives the following result.

[11] As an exercise, prove rigorously that T_A is linear.

Theorem 6.5
$\mathcal{L}(V, W)$ is a vector space.

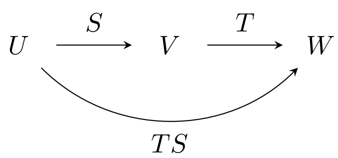
Additionally, we can show that the map $\mathcal{M} : \mathcal{L}(V, W) \to \mathbb{F}^{m,n}$ that sends T to $\mathcal{M}(T)$ is a linear map. To do this, for $S, T \in \mathcal{L}(V, W)$ and $c \in \mathbb{F}$, we need to show

- $\mathcal{M}(S + T) = \mathcal{M}(S) + \mathcal{M}(T)$,
- $\mathcal{M}(cT) = c \cdot \mathcal{M}(T)$.

The details of the proof are left as an exercise to the reader.

6.4 Matrix Multiplication

Before we define matrix multiplication, we will first discuss composition of linear maps. Suppose U, V, W are vector spaces and S, T are linear maps such that $S \in \mathcal{L}(U, V)$ and $T \in \mathcal{L}(V, W)$. Then, the composition of S and T is the map $TS \in \mathcal{L}(U, W)$:



The verification that TS is linear is left as an exercise.

Guiding Question
How do we express the composed map TS as a matrix?

Suppose $u_1, \ldots, u_k$, $v_1, \ldots, v_n$, and $w_1, \ldots, w_m$ are bases of U, V, and W respectively. Then, denote $\mathcal{M}(S)$ as the $n \times k$ matrix A and $\mathcal{M}(T)$ as the $m \times n$ matrix B.

Now, we will compute $\mathcal{M}(TS)$. For brevity, denote $\mathcal{M}(TS) = C$. First, we will find the first column of C. We have

$$\begin{aligned}
(TS)(u_1) &= T(S(u_1)) \\
&= T(A_{11}v_1 + \cdots + A_{n1}v_n) \\
&= A_{11}T(v_1) + \cdots + A_{n1}T(v_n) \\
&= A_{11}(B_{11}w_1 + \cdots + B_{m1}w_m) + \cdots + A_{n1}(B_{1n}w_1 + \cdots + B_{m1}w_n) \\
&= (A_{11}B_{11} + \cdots + A_{n1}B_{1n})w_1 + \cdots + (A_{11}B_{m1} + \cdots + A_{n1}B_{mn})w_m \\
&= \sum_{j=1}^{m} \left(\sum_{i=1}^{n} A_{i,1}B_{j,i} \right) w_j.
\end{aligned}$$

It follows that

$$C_{j1} = \sum_{i=1}^{n} A_{i,1}B_{j,i}.$$

Generalizing to the k^{th} column of C, we find that

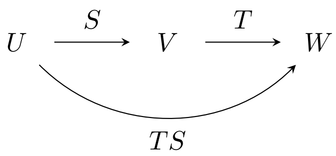
$$C_{jk} = \sum_{i=1}^{n} A_{i,k}B_{j,i}.$$

Thus, we now have defined matrix multiplication. In the next lecture, we will show the usefulness of this definition.

7 Matrix Multiplication (continued), Null Space, and Range

7.1 Review

Last time, we ended with the notion of composition of linear maps:



Note the order of the linear maps: applying TS is equivalent to first applying S, then T. Then, we computed the matrix of TS in terms of the matrices of S and T. Recall from the end of last lecture that for $\mathcal{M}(S) = A, \mathcal{M}(T) = B$, and $\mathcal{M}(TS) = C$, we have

$$C_{jk} = \sum_{i=1}^{n} A_{i,k} B_{j,i}.$$

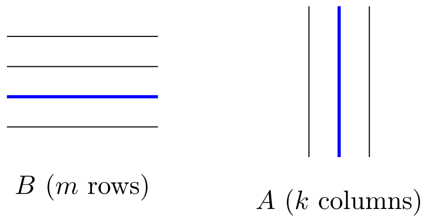
7.2 Matrix Multiplication (continued)

The above formula motivates the following definition.

Definition 7.1 (matrix multiplication)
Suppose B is an $m \times n$ matrix and A is an $n \times k$ matrix. The product BA is the matrix with entries

$$(BA)_{ij} = \sum_{r=1}^{k} B_{i,r} A_{r,j}.$$

Note that this is the same formula as in Section 7.1 with variables rearranged. To visualize this formula, we can interpret matrix B as consisting of m rows and matrix A consisting of k columns:



The product is only defined if the length of each column of B matches the length of each column of A. To find a specific entry in the product BA, we take the dot product of the i^{th} column of B and the j^{th} row of A. For example, taking the dot product of the two highlighted lines in the above diagram gives $(BA)_{3,2}$.

Our previous calculations give us the following result.

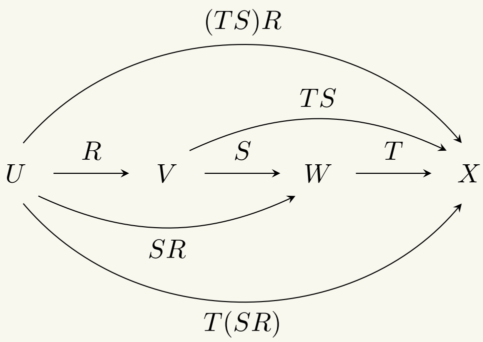
Theorem 7.2
Suppose $S \in \mathcal{L}(U, V)$ and $T \in \mathcal{L}(V, W)$. Then, $\mathcal{M}(TS) = \mathcal{M}(T)\mathcal{M}(S)$.

Note that the left-hand side uses the notion of composition of linear maps while the right-hand side uses matrix multiplication. In fact, our definition of matrix multiplication was constructed such that the above result holds. While you have likely seen the formula for matrix multiplication previously, here we have presented the motivation for defining matrix multiplication the way it is.

7.3 Properties of Matrix Multiplication

Whenever we have a property of composition of linear maps, there is an analogous property for matrix multiplication.

Fact 7.3 (Associativity)
Suppose U, V, W, and X are vector spaces and $R \in \mathcal{L}(U, V), S \in \mathcal{L}(V, W)$, and $T \in \mathcal{L}(W, X)$. The following picture



shows that $T(SR) = (TS)R$, so composition is associative.

Consequently, this result shows that matrix multiplication is associative. In particular, suppose A, B, and C are matrices such that BA and CB are defined.[a] Then, $C(BA) = (CB)A$.

[a] Recall that the product BA is defined when the number of columns in B is the same as the number of rows in A. The same requirement holds for CB.

Remark. *Suppose A and B are matrices. In general, matrix multiplication is* not *commutative (e.g. $AB \neq BA$). The following two reasons contribute to this fact:*

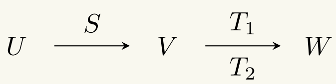
- *First, AB and BA may not both be defined or have di erent size. Suppose A is size $n \times k$ and B is size $m \times n$. Then, product BA is defined, but AB is only defined when $k = m$. Even if both products are defined (e.g. $k = m$), then AB is size $n \times n$ and BA is size $m \times m$, so the products will have di erent size unless $m = n$.*
- *Suppose AB and BA are both defined and have the same size, which implies that both A and B are $n \times n$ matrices. Then, it is still true in general that $AB \neq BA$. For instance, consider*

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

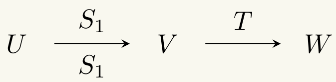
We calculate

$$AB = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = BA.$$

Fact 7.4 (Distributivity)
Suppose U, V, and W are vector spaces. Let $S \in \mathcal{L}(U, V)$ and $T_1, T_2 \in \mathcal{L}(W)$:



Then, $(T_1 + T_2)S = T_1S + T_2S$. Now, let $S_1, S_2 \in \mathcal{L}(U, V)$ and $T \in \mathcal{L}(W)$:



Then, $T(S_1 + S_2) = TS_1 + TS_2$. Thus, composition of linear maps is distributive, which implies that matrix multiplication is also distributive.

Consider the following special case. Let V be a vector space and consider the set of all linear maps from V to itself $\mathcal{L}(V, V)$, also denoted as $\mathcal{L}(V)$. Then, composition is a binary operation on $\mathcal{L}(V)$, since for $S, T \in \mathcal{L}(V)$, it follows that $TS \in \mathcal{L}(V)$. By Fact 7.3 and Fact 7.4, composition is associative and distributive. Additionally,

$\mathcal{L}(V)$ has an identity element id_V, which satisfies $\text{id}_V(v) = v$ for all $v \in V$ and $S = \text{id}_V S = S\text{id}_V$ for all $S \in \mathcal{L}(V)$.

This shows that $\mathcal{L}(V)$ is a *ring*, which is an algebraic structure equipped with addition and multiplication. Addition in $\mathcal{L}(V)$ is defined by the addition of linear maps, and multiplication is defined by the composition of linear maps.

Additionally, by choosing any basis of V, we can form a bijection between $\mathcal{L}(V)$ and $\mathbb{F}^{n,n}$. Thus, it follows that $\mathbb{F}^{n,n}$ is also a ring, where multiplication is defined by matrix multiplication.

Remark. *When working with matrices, it is important to di erentiate* scalar multiplication *and* matrix multiplication. *Scalar multiplication is when you multiply a matrix by a number c, resulting in a matrix of the same size. Matrix multiplication involves multiplying two matrices BA and the product has the same number of columns as A and the same number of rows as B.*

However, scalar multiplication is actually a special case of matrix multiplication. For any $c \in \mathbb{F}$, suppose we wish to find

$$c\begin{pmatrix}1 & 2 & 3\\ 4 & 5 & 6\end{pmatrix} = \begin{pmatrix}c & 2c & 3c\\ 4c & 5c & 6c\end{pmatrix} = \begin{pmatrix}? & ?\\ ? & ?\end{pmatrix}\begin{pmatrix}1 & 2 & 3\\ 4 & 5 & 6\end{pmatrix}.$$

Let A be the matrix on the left-hand side. We can see that A is the matrix of a linear map from $\mathbb{F}^3$ to $\mathbb{F}^2$. Then, to multiply by the scalar c, we must apply $c \cdot \text{id}_{\mathbb{F}^2}$, which gives us the composition

$$\mathbb{F}^3 \xrightarrow{A} \mathbb{F}^2 \xrightarrow{c\cdot\text{id}_{\mathbb{F}^2}} \mathbb{F}^2.$$

Thus, the matrix on the right-hand side should be $\mathcal{M}(c \cdot \text{id}_{\mathbb{F}^2})$, so

$$c\begin{pmatrix}1 & 2 & 3\\ 4 & 5 & 6\end{pmatrix} = \begin{pmatrix}c & 2c & 3c\\ 4c & 5c & 6c\end{pmatrix} = \begin{pmatrix}c & 0\\ 0 & c\end{pmatrix}\begin{pmatrix}1 & 2 & 3\\ 4 & 5 & 6\end{pmatrix}.$$

For this reason, we denote scalar matrices *as matrices of the form*

$$\begin{pmatrix}c & & 0\\ & \ddots & \\ 0 & & c\end{pmatrix}$$

for any $c \in \mathbb{F}$, and these matrices are equivalent to $\mathcal{M}(c \cdot \text{id}_V)$ for any vector space V.

Lastly, there is another way to interpret scalar multiplication as a special case of matrix multiplication, which will be left as an exercise to figure out (Hint: in the above example, we found a matrix to left-multiply with A; what if we try to find a matrix to right-multiply with A?).

7.4 Meaning of Ad

Suppose A is a $m\times n$ matrix and d is a $n\times 1$ column vector. Thus, the product Ad is a $m\times 1$ column vector. Now, we will explain the meaning of Ad in terms of linear maps.

Let $A = \mathcal{M}(T)$ for some linear map $T \in \mathcal{L}(V, W)$ with respect to bases $v_1, \ldots, v_n$ of V and $w_1, \ldots, w_m$ of W. Let d_i denote the i^{th} entry of d. Define $v = d_1v_1 + \cdots + d_nv_n$. We wish to write $T(v)$ as a linear combination of $w_1, \ldots, w_m$. First, we can express this as

$$T(v) = d_1T(v_1) + \cdots + d_nT(v_n) = \sum_{i=1}^{n} d_iT(v_i).$$

Also, since $A = \mathcal{M}(T)$, we know that

$$T(v_i) = A_{1i}w_1 + \cdots + A_{mi}w_m = \sum_{j=1}^{m} A_{ji}w_j.$$

Then, we have

$$\begin{aligned} T(v) &= \sum_{i=1}^{n} d_i T(v_i) \\ &= \sum_{i=1}^{n} d_i \left(\sum_{j=1}^{m} A_{ji} w_j \right) \\ &= \sum_{j=1}^{m} \left(\sum_{i=1}^{n} A_{ji} d_i \right) w_j. \end{aligned}$$

Thus, we have expressed $T(v)$ as a linear combination of $w_1, \ldots, w_m$.

Now, let $(Ad)_j$ denote the j^{th} entry of Ad. By the formula for matrix multiplication,

$$(Ad)_j = \sum_{i=1}^{n} A_{ji} d_i,$$

which is the same expression that we got for the coefficient of the w_j term in $T(v)$. Therefore, Ad tells us the "coordinates" of $T(v)$ with respect to the basis $w_1, \ldots, w_m$.

What is the meaning of multiplying a row vector with a matrix? For instance, suppose r is a $1 \times m$ row vector and A is a $m \times n$ matrix. Then, rA is a $1 \times n$ row vector. The meaning of rA is not as straightforward, but we will come back to the meaning of this row vector later in this course.

7.5 Null Space

Now, we will shift our attention away from matrices and back towards linear maps, starting with null spaces.

Definition 7.5 (null space)
Suppose $T \in \mathcal{L}(V, W)$. The **null space** of T, denoted $\text{Null}(T)$, is defined as

$$\text{Null}(T) = \{v \in V \mid T(v) = 0\}.$$

We call it the "null space" because it is actually a subspace.

Proposition 7.6
Suppose $T \in \mathcal{L}(V, W)$. Then, $\text{Null}(T)$ is a subspace of V.

Proof. Because T is a linear map, $T(\vec{0}) = \vec{0}$, so $\vec{0} \in \text{Null}(T)$. Suppose $v_1, v_2 \in \text{Null}(T)$. Then,

$$T(v_1 + v_2) = T(v_1) + T(v_2) = \vec{0} + \vec{0} = \vec{0},$$

so $v_1 + v_2 \in \text{Null}(T)$. Thus, $\text{Null}(T)$ is closed under addition. Now, suppose $v \in \text{Null}(T)$ and $c \in \mathbb{F}$. Then,

$$T(cv) = cT(v) = c\vec{0} = \vec{0},$$

so $cv \in \text{Null}(T)$. Thus, $\text{Null}(T)$ is closed under scalar multiplication. Therefore, $\text{Null}(T)$ is a subspace. □

Additionally, we call $\dim \text{Null}(T)$ the ***nullity*** of T.

Definition 7.7 (injective)
A map $T : V \to W$ is **injective** if $Tu = Tv$ implies $u = v$.

If T is injective, we could rephrase the above definition as $u \neq v$ implies $Tu \neq Tv$. In other words, if T is injective, T maps distinct inputs to distinct outputs.

The above definition is related to null spaces by the following result.

Theorem 7.8
Suppose $T \in \mathcal{L}(V, W)$. Then, $\text{Null}(T) = \{\vec{0}\}$ if and only if T is injective.

Proof. First, we will prove the forward direction. Suppose $\text{Null}(T) = \{\vec{0}\}$ and $v_1, v_2 \in V$ such that $T(v_1) = T(v_2)$. Then,

$$T(v_1 - v_2) = T(v_1) - T(v_2) = \vec{0},$$

so $v_1 - v_2 \in \text{Null}(T)$. This implies that $v_1 - v_2 = \vec{0}$, so $v_1 = v_2$. Therefore, T is injective.

Now, we will prove the backwards direction. Suppose T is injective. Since T is a linear map, we know that $T(\vec{0}) = \vec{0}$, so $\vec{0} \in \text{Null}(T)$. Because T is injective, it follows that no other vectors in V can map to $\vec{0}$. Therefore, $\text{Null}(T) = \{\vec{0}\}$. □

The above result is quite interesting; the statement $\text{Null}(T) = \{\vec{0}\}$ says that only the zero vector in V can map to the zero vector in W. From this, we can deduce that for any $v_1, v_2 \in V$ with $v_1 \neq v_2$, it follows that $Tv_1 \neq Tv_2$.

7.6 Range

Now, we will discuss the range of a linear map, which has many parallels to the null space.

Definition 7.9 (range)
Suppose $T \in \mathcal{L}(V, W)$. The **range** of T, denoted $\text{Range}(T)$, is defined as

$$\text{Range}(T) = \{T(v) \mid v \in V\}.$$

The range of T is also sometimes called the ***image*** of T. Note that while $\text{Null}(T)$ is a subset of V, $\text{Range}(T)$ is a subset of W.

Proposition 7.10
Suppose $T \in \mathcal{L}(V, W)$. Then, $\text{Range}(T)$ is a subspace of W.

Proof. The proof is similar to the proof of Proposition 7.6 and is left as an exercise to the reader. □

Additionally, we call $\dim \text{Range}(T)$ the ***rank*** of T.

Definition 7.11 (surjective)
A map $T : V \to W$ is **surjective** if its range equals W.

From the above definition, it is clear that $\text{Range}(T) = W$ if and only if T is surjective.

Example 7.12
Suppose $T \in \mathcal{L}(V, W)$. Consider the special case where $\text{Range}(T) = \{\vec{0}\}$. This implies that $Tv = \vec{0}$ for any $v \in V$, so T is the zero map.

Now, consider the special case where $\text{Null}(T) = V$. This also implies that $Tv = \vec{0}$ for any $v \in V$, so T is the zero map.

Example 7.13
Consider the linear map $D : \mathcal{P}_m(\mathbb{F}) \to \mathcal{P}_m(\mathbb{F})$ defined by $Dp = p'$ for all polynomials $p \in \mathcal{P}_m(\mathbb{F})$. Then,

$$\operatorname{Null}(D) = \{\text{constant polynomials}\} = \mathcal{P}_0(\mathbb{F}),$$

so the nullity of D is $\dim \mathcal{P}_0(\mathbb{F}) = 1$.

Now, to find $\operatorname{Range}(D)$, note that $\operatorname{Range}(D) \subset \mathcal{P}_{m-1}(\mathbb{F})$. Also, note that $D(\frac{1}{i+1}x^{i+1}) = x^i$, so $1, x, x^2, \dots, x^{m-1} \in \operatorname{Range}(D)$. Therefore, $\operatorname{span}(1, x, x^2, \dots, x^{m-1}) = \mathcal{P}_{m-1}(\mathbb{F}) \subset \operatorname{Range}(D)$, so $\operatorname{Range}(D) = \mathcal{P}_{m-1}(\mathbb{F})$. It follows that the rank of D is $\dim \mathcal{P}_{m-1}(\mathbb{F}) = m$.

Example 7.14
Suppose $T \in \mathcal{L}(\mathbb{F}^3, \mathbb{F}^3)$ such that $\mathcal{M}(T)$ is

$$A = \begin{pmatrix} 1 & 3 & 1 \\ 2 & 6 & 1 \\ 0 & 0 & 2 \end{pmatrix}.$$

Then, $\operatorname{Range}(T)$ is the span of the columns of A, which we can calculate to be

$$\operatorname{Range}(T) = \operatorname{span}\left(\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} \right).$$

Thus, the rank of T is 2.

To find $\operatorname{Null}(T)$, we must solve the system of equations

$$\begin{pmatrix} 1 & 3 & 1 \\ 2 & 6 & 1 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

In this example, we can calculate

$$\operatorname{Null}(T) = \operatorname{span}\left(\begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} \right),$$

so the nullity of T is 1.

8 Rank-Nullity Theorem, Isomorphisms, Product Space, and Dual Space

8.1 Review

Last time, we introduced the null space and range of a linear map (Definitions 7.5 and 7.9). For $\mathcal{L}(V, W)$, recall that $\text{Null}(T)$ is a subset of V and contains all vectors that T sends to $\vec{0}$, while $\text{Range}(T)$ is a subset of W and contains all vectors in the image of T

8.2 Rank-Nullity Theorem

The next result is one of the most important theorems in linear algebra.

Theorem 8.1 (Rank-Nullity theorem)
Suppose V is finite-dimensional and $T \in \mathcal{L}(V, W)$. Then,

$$\dim \text{Null}(T) + \dim \text{Range}(T) = \dim V.$$

Proof. Let $V_1, \ldots, v_k$ be a basis of $\text{Null}(T)$. Since this list is linearly independent, it can be extended to a basis $v_1, \ldots, v_n$ of V. This implies that $\dim \text{Null}(T) = k$ and $\dim V = n$.

Now, we will show that $T(v_{k+1}), \ldots, T(v_n)$ is a basis of $\text{Range}(T)$. First, we will show that this list is linearly independent. Suppose $c_{k+1}, \ldots, c_n \in \mathbb{F}$ such that $c_{k+1}T(v_{k+1}) + \cdots + c_nT(v_n) = \vec{0}$. It follows that

$$T(c_{k+1}v_{k+1} + \cdots + c_nv_n) = \vec{0},$$

so $c_{k+1}v_{k+1} + \cdots + c_nv_n \in \text{Null}(T)$. Because $v_1, \ldots, v_k$ is a basis of $\text{Null}(T)$, there exist $c_1, \ldots, c_k \in \mathbb{F}$ such that

$$c_1v_1 + \cdots + c_kv_k = c_{k+1}v_{k+1} + \cdots + c_nv_n.$$

Thus, $c_1v_1 + \cdots + c_kv_k - c_{k+1}v_{k+1} - \cdots - c_nv_n$. Since $v_1, \ldots, v_n$ is a basis, it follows that $c_1, \ldots, c_n$ are all 0. In particular, $c_{k+1}, \ldots, c_n$ are all 0, so $T(v_{k+1}), \ldots, T(v_n)$ is linearly independent.

To show that $T(v_{k+1}), \ldots, T(v_n)$ span $\text{Range}(T)$, let $w \in \text{Range}(T)$. This implies that $w = T(v)$ for some $v \in V$. Since $v_1, \ldots, v_n$ is a basis, we can write $v = c_1v_1 + \cdots + c_nv_n$. Then,

$$\begin{aligned}
w &= T(v) \\
&= T(c_1v_1 + \cdots + c_nv_n) \\
&= c_1T(v_1) + \cdots + c_kT(v_k) + c_{k+1}T(v_{k+1}) + \cdots + c_nT(v_n) \\
&= c_{k+1}T(v_{k+1}) + \cdots + c_nT(v_n)
\end{aligned}$$

because $v_1, \ldots, v_k \in \text{Null}(T)$, so $c_1T(v_1) + \cdots + c_kT(v_k) = \vec{0}$. Thus, $T(v_{k+1}), \ldots, T(v_n)$ span $\text{Range}(T)$.

Therefore, $T(v_{k+1}), \ldots, T(v_n)$ is a basis of $\text{Range}(T)$. It follows that $\dim \text{Range}(T) = n - k$, so

$$\dim \text{Null}(T) + \dim \text{Range}(T) = k + (n - k) = n = \dim V,$$

as desired. □

The above result is very useful: given that we know $\dim V$, we only need to compute one of $\dim \text{Null}(T)$ or $\dim \text{Range}(T)$, and Theorem 8.1 will tell us the other. For instance, suppose $T \in \mathcal{L}(V, W)$ and consider the following cases:

- T is surjective: it follows that $\text{Range}(T) = W$, so $\dim \text{Null}(T) = \dim V - \dim W$.
- $T$is injective: it follows that $\text{Null}(T) = \{0\}$, so $\dim \text{Range}(T) = \dim V$.

8.3 Invertible Linear Maps

We will begin this section by defining invertible linear maps.

Definition 8.2 (invertible, inverse)
A linear map $T \in \mathcal{L}(V, W)$ is called **invertible** if there exists a linear map $S \in \mathcal{L}(W, V)$ such that $ST = \mathrm{id}_V$ and $TS = \mathrm{id}_W$.

Additionally, S is unique and called the **inverse** of T, denoted T^{-1}.

While the above definition is quite complicated, we often think of invertible linear maps in a simpler way.

Definition 8.3 (isomorphism)
A linear map $T \in \mathcal{L}(V, W)$ is an **isomorphism** if it is both injective and surjective.

The next result relates the two definitions.

Theorem 8.4
Suppose $T \in \mathcal{L}(V, W)$. Then, T is invertible if and only if it is an isomorphism.

Proof. First, suppose T is invertible. We wish to show that T is an isomorphism. To show that T is injective, let $v \in V$ such that $T(v) = 0$. Let S be an inverse of T. It follows that

$$ST(v) = S(\vec{0}) = \vec{0}.$$

However, we know that $ST = \mathrm{id}_V$, so $ST(v) = v$. Thus, $v = \vec{0}$, so $\mathrm{Null}(T) = \{\vec{0}\}$ and T is injective. To show that T is surjective, let $w \in W$. Since $TS = \mathrm{id}_W$, it follows that $T(S(w)) = w$. Thus, $w \in \mathrm{Range}(T)$, which implies that $\mathrm{Range}(T) = W$, so T is surjective. Therefore, T is an isomorphism.

Now, suppose T is an isomorphism. We wish to show that T is invertible. To do this, we will construct an inverse $S \in \mathcal{L}(W, V)$. Let $w_1, \dots, w_n$ be a basis of W. Since T injective and surjective, there exists a unique $v_i \in V$ such that $T(v_i) = w_i$. Define $S(w_i) = v_i$ for $i = 1, \dots, n$. Because $w_1, \dots, w_n$ is a basis, this uniquely determines the linear map S. It is clear that

$$T(S(w_i)) = T(v_i) = w_i,$$

so $TS = \mathrm{id}_W$. To prove that $ST = \mathrm{id}_V$, note that

$$T(ST) = (TS)T = \mathrm{id}_W T = T.$$

This implies that $T(ST(v)) = T(v)$ for any $v \in V$, so $ST(v) = v$ because T is injective. Thus, $ST = \mathrm{id}_V$. Therefore, S is an inverse of T, so T is invertible. □

We have shown that invertible linear maps and isomorphisms are equivalent, so we will use the two terms interchangeably from now on.

In Definition 8.2, we claimed that invertible linear maps have a unique inverse. The proof of this statement is left as an exercise.

8.4 Isomorphic Vector Spaces

The next definition is closely related to the concepts in the previous section.

Definition 8.5 (isomorphic)
Two vector spaces V and W are called **isomorphic** if there exists an isomorphism $T \in \mathcal{L}(V, W)$.

The notation $V \cong W$ or $V \simeq W$ are sometimes used to denote isomorphic vector spaces. Additionally, note that it is possible for V and W to be isomorphic by multiple different isomorphisms. For instance, $T \in \mathcal{L}(\mathbb{R}, \mathbb{R})$

defined by $T(v) = cv$ is an isomorphism for any nonzero scalar c. Of course, this example is showing that $\mathbb{R}$ is isomorphic to itself, which is not very noteworthy.

Consider the following examples of isomorphisms and isomorphic vector spaces.

Example 8.6
$\mathcal{L}(\mathbb{F}^m, \mathbb{F}^n)$ and $\mathbb{F}^{m,n}$ are isomorphic. We proved in a previous lecture that the linear map between $\mathcal{L}(\mathbb{F}^m, \mathbb{F}^n)$ and $\mathbb{F}^{n,m}$ is both injective and surjective, which is precisely the definition of an isomorphism.

Example 8.7
Suppose V is a vector space. Let $v_1, \ldots, v_n$ be a basis of V and let $T \in \mathcal{L}(\mathbb{F}^n, V)$ defined by

$$T((c_1, \ldots, c_n)) = \sum_{i=1}^{n} c_i v_i.$$

We can see that T is injective if and only if $v_1, \ldots, v_n$ is linearly independent. Similarly, T is surjective if and only if $v_1, \ldots, v_n$ span V. Since $v_1, \ldots, v_n$ is a basis, it follows that T is both injective and surjective, so $\mathbb{F}^n$ and V are isomorphic. This example shows that for any vector space V, there exists an isomorphism from $\mathbb{F}^{\dim V}$ to V.

The next result gives an easy way to tell if two vector spaces are isomorphic.

Proposition 8.8
Two finite-dimensional vector spaces V and W are isomorphic if and only if $\dim V = \dim W$.

Proof. First, suppose V and W are isomorphic. Then, there exists an isomorphism $T \in \mathcal{L}(V, W)$. Let $v_1, \ldots, v_n$ be a basis of V. We will show that $T(v_1), \ldots, T(v_n)$ is a basis of W. Because T is injective, it can be shown that $T(v_1), \ldots, T(v_n)$ is linearly independent. Similarly, because T is surjective, it can be shown that $T(v_1), \ldots, T(v_n)$ spans W. Thus, $T(v_1), \ldots, T(v_n)$ is a basis of W, so $\dim V = n = \dim W$.

Now, suppose $\dim V = \dim W$. Let $v_1, \ldots, v_n$ and $w_1, \ldots, w_n$ be bases of V and W, respectively. Let $T \in \mathcal{L}(V, W)$ be defined by $T(v_i) = w_i$ for $i = 1, \ldots, n$. It can be shown that T is an isomorphism (the details of the proof are left as an exercise), so V and W are isomorphic. □

For instance, we showed that $\mathcal{L}(\mathbb{F}^m, \mathbb{F}^n)$ and $\mathbb{F}^{n,m}$ are isomorphic in Example 8.6. The vector space $\mathcal{L}(\mathbb{F}^m, \mathbb{F}^n)$ is quite abstract, so we may not be able to figure out the dimension directly. However, we know $\dim \mathbb{F}^{n,m} = mn$, which implies that $\dim \mathcal{L}(\mathbb{F}^m, \mathbb{F}^n) = mn$ by Proposition 8.8.

8.5 Product Space

The product of vector spaces introduces the notion of combining smaller vector spaces to create a larger vector space.

Definition 8.9 (product of vector spaces)
Suppose V_1 and V_2 are vector spaces over $\mathbb{F}$.

- The **product** $V_1 \times V_2$ is defined by
$$V_1 \times V_2 = \{(v_1, v_2) \mid v_1 \in V_1, v_2 \in V_2\}.$$
- Addition on $V_1 \times V_2$ is defined by
$$(u_1, u_2) + (v_1, v_2) = (u_1 + v_1, u_2 + v_2).$$
- Scalar multiplication on $V_1 \times V_2$ is defined by
$$c(v_1, v_2) = (cv_1, cv_2).$$

The next result shows that the product of vector spaces is a vector space.

Proposition 8.10
Suppose V_1 and V_2 are vector spaces over $\mathbb{F}$. Then, $V_1 \times V_2$ is a vector space over $\mathbb{F}$.

Proof. The proof is left as an exercise to the reader. □

The product of vector spaces can be similarly defined for any finite number of vector spaces $V_1, \dots, V_m$, defined by
$$V_1 \times \cdots \times V_m = \{(v_1, \dots, v_m) \mid v_1 \in V_1, \dots, v_m \in V_m\}.$$
It can be similarly shown that $V_1 \times \cdots \times V_m$ is a vector space.

Recall the definition of a direct sum (Definition 3.3). We will show how the product of vector spaces is related to direct sum. Suppose $U_1, \dots, U_m$ are subspaces of vector space V. Consider the map $T : U_1 \times \cdot \times U_m \to U_1 + \cdots + U_m$ defined by
$$T((u_1, \dots, u_m)) = u_1 + \cdots + u_m.$$
It is easy to show that T is linear. Now, it follows that if $U_1 \oplus \cdots \oplus U_m$ is a direct sum, then T is an isomorphism. To prove this, note that by the definition of direct sum, each vector in $U_1 + \cdots + U_m$ can be expressed uniquely as a sum $u_1 + \cdots + u_m$, which implies that T is injective and surjective.

Thus, $U_1 \times \cdot \times U_m$ and $U_1 \oplus \cdots \oplus U_m$ are isomorphic. This fact gives intution to the following result, which we will not prove.

Proposition 8.11
Suppose $U_1, \dots, U_m$ are subspaces of V. Then, $U_1 + \cdots + U_m$ is a direct sum if and only if
$$\dim U_1 + \cdots + \dim U_m = \dim(U_1 + \cdots + U_m).$$

The above result relies on the fact that
$$\dim(U_1 \times \cdots \times U_m) = \dim U_1 + \cdots + \dim U_m.$$

To gain some intuition on why this holds, consider $\mathbb{F}^m \times \mathbb{F}^n$. The linear map that sends $((x_1, \dots, x_m), (y_1, \dots, y_n)) \in \mathbb{F}^m \times \mathbb{F}^n$ to $(x_1, \dots, x_m, y_1, \dots, y_n) \in \mathbb{F}^{m+n}$ is clearly an isomorphism, so $\mathbb{F}^m \times \mathbb{F}^n \cong \mathbb{F}^{m+n}$. Thus, $\dim(\mathbb{F}^m \times \mathbb{F}^n) = \dim \mathbb{F}^{m+n} = m + n = \dim \mathbb{F}^m + \dim \mathbb{F}^n$.

8.6 Dual Space

The dual space introduces a new vector space closely related to the original vector space.

Definition 8.12 (dual space)
Suppose V is a vector space. The **dual space** of V, denoted V', is defined as $\mathcal{L}(V, \mathbb{F})$.

A linear map from V to $\mathbb{F}$ is called a *linear functional* on V. Thus, the dual space can also be defined as the set of all linear functionals on V.

Remark. *As a sidenote, consider the geometric meaning of a linear functional. We are projecting a high-dimensional vector space V onto a one-dimensional vector space and measuring its length. In other words, we are measuring the "shadow" of V in a certain direction.*

Consider the following example of linear functionals.

Example 8.13
Suppose $\varphi : \mathcal{P}_m(\mathbb{F}) \to \mathbb{F}$. All of the following are linear functionals:

- $\varphi(p) = p(1)$
- $\varphi(p) = p'(2) + 2p''(3)$
- $\varphi(p) = \int_0^1 p(x)\, dx$.

To better understand the dual space, consider the dual space of $\mathbb{F}^n$ as an example. We have

$$(\mathbb{F}^n)' = \mathcal{L}(\mathbb{F}^n, \mathbb{F}) \cong \mathbb{F}^{1,n}.$$

Note that $\mathbb{F}^n$ and $\mathbb{F}^{1,n}$ both have dimension n, but the elements of $\mathbb{F}^n$ are typically column vectors while the elements of $\mathbb{F}^{1,n}$ are row vectors.

More concretely, suppose $\varphi : \mathbb{F}^n \to \mathbb{F}$ is a linear functional. Let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{F}^n$. Then, φ is uniquely determined by the n numbers $\varphi(e_1), \ldots, \varphi(e_n)$. This gives a bijection from $\mathcal{L}(\mathbb{F}^n, \mathbb{F})$ to $\mathbb{F}^n$ defined by sending $\varphi \in \mathcal{L}(\mathbb{F}\cdot\mathbb{F})$ to $(\varphi(e_1), \ldots, \varphi(e_n)) \in \mathbb{F}^n$. Intuitively, this makes sense because we showed earlier that $\mathbb{F}^n$ and $\mathcal{L}(\mathbb{F}^n, \mathbb{F})$ are both n-element vectors, only differing in being column or row vectors.

9 Dual Maps and Gaussian Elimination

9.1 Review

Last time, we introduced the notion of dual space. Recall that if V is a vector space over $\mathbb{F}$, the dual space of V is defined as $V' = \mathcal{L}(V, \mathbb{F})$.

9.2 Dual Basis

We will continue our discussion of the dual space by introducing the notion of a dual basis. The idea of a dual basis is prompted by the following question.

Guiding Question
Given a basis $v_1, \dots, v_n$ of V, how do we construct a basis of V'?

Note that a basis of V' would be made up of linear functionals $\varphi_i : V \to \mathbb{F}$. We define $\varphi_i \in V'$ by

$$\varphi_i(v_j) = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases}.$$

The list of linear functionals $\varphi_1, \dots, \varphi_n$ is called the ***dual basis*** of V'.

Lemma 9.1
The dual basis is a basis of V'.

Proof. First, we will show that the dual basis is linearly independent. Suppose there exists scalars $c_1, \dots, c_n$ such that $c_1\varphi_1 + \cdots + c_n\varphi_n = 0$. We compute that

$$(c_1\varphi_1 + \cdots + c_n\varphi_n)(v_j) = c_1\varphi(v_j) + \cdots + c_n\varphi(v_j) = c_j$$

for all $j = 1, \dots, n$. However, we also compute

$$(c_1\varphi_1 + \cdots + c_n\varphi_n)(v_j) = 0(v_j) = 0,$$

so $c_1 = \cdots = c_n = 0$. Thus, $\varphi_1, \dots, \varphi_n$ is linearly independent.

Now, we will show that the dual basis spans V'. Suppose $\varphi \in V'$ and $v_1, \dots, v_n$ is a basis of V. Define $c_i = \varphi(v_i)$ for all $i = 1, \dots, n$. We will show that $\varphi = c_1\varphi_i + \cdots + c_n\varphi_n$. To prove this, we will show that $\varphi(v_j) = (c_1\varphi_1 + \cdots + c_n\varphi_n)(v_j)$ for all $j = 1, \dots, n$. We compute

$$(c_1\varphi_1 + \cdots + c_n\varphi_n)(v_j) = c_1\varphi(v_j) + \cdots + c_n\varphi(v_j) = c_j,$$

which is equal to $\varphi(v_j)$ by definition. Since φ was an arbitrary element in V', it follows that $\varphi_1, \dots, \varphi_n$ span V'.

Therefore, the dual basis is a basis of V'. □

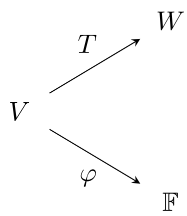
Note that the definition of the dual basis follows relatively intuitively from the given basis $v_1, \dots, v_n$ of V. Furthermore, an important consequence of Lemma 9.1 is that $\dim V = \dim V'$.

9.3 Dual Map

Now, we will connect the notion of dual spaces with linear maps. The idea of a dual map is prompted by the following question.

Guiding Question
Suppose T is a linear map from V to W. How would you construct a "dualized" version of T between the dual spaces V' and W'?

First, recall that an element $\varphi \in V'$ is a linear functional from V to $\mathbb{F}$:



We wish to construct a map from W to $\mathbb{F}$, which would be an element of W'. Given the two maps T and φ, it is not possible to construct such a map.[12]

On the other hand, suppose we are given a linear functional $\varphi \in W'$:

$$V \xrightarrow{\quad T \quad} W \xrightarrow{\quad \varphi \quad} \mathbb{F}$$

This time, we wish to construct a map from V to $\mathbb{F}$, which would be an element of V'. This can be done by the composition $\varphi T \in V'$. This procedure creates the ***dual map*** of T, denoted T', where $T' \in \mathcal{L}(W', V')$ is defined by $T'(\varphi) = \varphi T$ for $\varphi \in W'$.[13]

Example 9.2
Suppose $T \in \mathcal{L}(\mathbb{F}, \mathbb{F}^3)$ defined by $T(1) = (6, 5, 4)$. We wish to find the dual map of T.

Note that $T' \in \mathcal{L}((\mathbb{F}^3)', (\mathbb{F})')$. For brevity, let $V = \mathbb{F}$ and $W = \mathbb{F}^3$. First, we must find the dual spaces of V and W Let e and (e_1, e_2, e_3) be the standard bases of V and W, respectively, and let φ and $(\varphi_1, \varphi_2, \varphi_3)$ be the dual bases of V' and W', respectively.

Now, we can write the matrix of T' with respect to the dual bases. To do this, we must compute $T'(\varphi_1), T'(\varphi_2)$, and $T'(\varphi_3)$. We calculate

$$(T'(\varphi_1))(e) = \varphi_1 T(e) = \varphi_1((6e_1 + 5e_2 + 4e_3)) = 6 = 6\varphi(e),$$

which implies that $T'(\varphi_1) = 6\varphi$. Similarly, we can calculate $T'(\varphi_2) = 5\varphi$ and $T'(\varphi_3) = 4\varphi$. Thus, the matrix of T' is

$$\mathcal{M}(T') = \begin{pmatrix} 6 & 5 & 4 \end{pmatrix}.$$

We can also compute

$$\mathcal{M}(T) = \begin{pmatrix} 6 \\ 5 \\ 4 \end{pmatrix},$$

so $\mathcal{M}(T')$ is the transpose of $\mathcal{M}(T)$.

The observation we made $\mathcal{M}(T') = (\mathcal{M}(T))^t$ in the above example is actually a general phenomenon.[14] Rigorously, suppose $T \in \mathcal{L}(V, W)$ and $\mathcal{M}(T)$ is the matrix of T with respect to bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$ of V and W, respectively. Let $\mathcal{M}(T')$ be the matrix of T' with respect to the dual bases $\varphi_1, \ldots, \varphi_n$ and $\psi_1, \ldots, \psi_m$ of V' and W', respectively, Then, it follows that $\mathcal{M}(T)$ and $\mathcal{M}(T')$ are transposes of each other. We will not give a rigorous proof, but Example 9.2 gives much of the intuition needed.

Additionally, a ***symmetric matrix*** is a square matrix A whose entries are determined by $A_{i,j} = A_{j,i}$. These conditions imply that $A^t = A$. Thus, if T is a linear map such that $\mathcal{M}(T)$ is symmetric, it follows that $\mathcal{M}(T) = \mathcal{M}(T')$.

Thus, we now have shown how to "dualize" vector spaces, bases, and linear maps.

9.4 Applications of Gaussian Elimination

At its core, linear algebra is about solving linear equations. In this section, we will discuss Gaussian elimination, a computational tool to help us achieve this goal.

[12] Note that if T is an isomorphism, we could construct φT^{-1} as a linear map from W to $\mathbb{F}$, but this does not work for general T.
[13] An important observation is that for $T \in \mathcal{L}(V, W)$, the dual map $T' \in \mathcal{L}(W', V')$ has the order of V and W reversed.
[14] Recall that the *transpose* of a matrix A, denoted A^t, is the matrix obtained from A by interchanging the rows and columns. In other words, if A is a $m \times n$ matrix, A^t is the $n \times m$ matrix whose entries are given by $(A^t)_{i,j} = A_{j,i}$.

Consider the system of equations

$$\begin{aligned} x_1 + 2x_2 + 5x_3 &= 0 \\ -x_1 - 2x_2 - 2x_3 &= 0 \\ x_1 + x_2 &= 0. \end{aligned}$$

This is an example a system of ***homogeneous*** linear equations because the right-hand sides are all 0. If any of the right-hand sides were not 0, these would be ***inhomogeneous*** linear equations. For homogeneous equations, there is always a trivial solution in which all variables are set to 0, so our goal is to find all nontrivial solutions, if any. For inhomogeneous equations, our goal is to find if any solutions exist at all, and if they do, find all solutions.

We can also rewrite system of equations in terms of matrices, which we will use to define homogeneous and inhomogeneous linear equations more concretely. First, we will discuss homogeneous equations.

Definition 9.3 (homogeneous)
A system of m **homogeneous** linear equations in variables $x_1, \ldots, x_n$ can be written as

$$Ax = \vec{0}$$

where $(x_1, \ldots, x_n)$ is written as a column vector and A is an $m \times n$ matrix.

For instance, the system of equations at the start of this section can be rewritten in matrix form as

$$\begin{pmatrix} 1 & 2 & 5 \\ -1 & -2 & -2 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

which is of the form $Ax = \vec{0}$. The importance of writing this in matrix is form is that A can be viewed as the matrix form of a linear map $T \in \mathcal{L}(\mathbb{F}^3, \mathbb{F}^3)$. Multiplying the product Ax is equivalent to applying T to the vector $x \in \mathbb{F}^3$ Thus, solving $Ax = \vec{0}$ is equivalent to finding all x such that $Tx = \vec{0}$, which occurs when $x \in \text{Null}(T)$. Therefore, finding solutions to a homogeneous system of equations is no more than finding the null space of a linear map. Gaussian elimination will help us find a basis of $\text{Null}(T)$, which corresponds to the fundamental solutions of homogeneous linear equations.[15]

Now, let us consider inhomogeneous linear equations.

Definition 9.4 (inhomogeneous)
A system of m **inhomogeneous** linear equations in variables $x_1, \ldots, x_n$ can be written as

$$Ax = \vec{0}$$

where $(x_1, \ldots, x_n)$ is written as a column vector, A is an $m \times n$ matrix, and $b \in \mathbb{F}^m$ is a column vector.

For now, we will only consider whether $Ax = b$ has a solution or not. Since A is the matrix form of a linear map T, the system of equations $Ax = b$ having a solution is equivalent to $b \in \text{Range}(T)$. Once again, Gaussian elimination will help us determine whether or not a solution exists.

9.5 Reduced Row Echelon Form

Now that we understand the usefulness of Gaussian elimination, will describe how it works. Gaussian elimination is an algorithm that starts with an $m \times n$ matrix A and performs a series of ***elementary row operations***, outputting a matrix of the same size in ***reduced row echelon form***. In this section, we will define reduced row echelon form and demonstrate why it is important.

Informally, a matrix is in reduced row echelon form if it looks like:

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

[15] A general solution can be written uniquely as a linear combination of fundamental solutions.

where the entries marked by $*$ are arbitrary (can be zero or nonzero).

Remark. *Note that it is not necessary for a matrix in reduced row echelon form to have a 1 in the first column. Also, it is possible for it to have rows consisting of only zeroes at the bottom. For instance, the matrix*

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

is in reduced row echelon form.

The entries containing leading 1's are called the ***pivots***; in other words, the pivot in i^{th} row is the first nonzero entry in the row. Additionally, if a matrix has r pivots, then the pivots must lie in the topmost r rows. Furthermore, each pivot must lie to the right of the pivot in the row above.

A matrix is in reduced row echelon form is it satisfies:

1. All entries to the left of a pivot are 0.
2. All pivots are 1.
3. All entries above a pivot are 0.
4. All rows below the pivot rows must consist only of zeros.

Formally, suppose E is an $m \times n$ matrix with pivots as the $(1, j_1), \ldots, (r, j_r)$ entries. Then, $0 \leq r \leq m$ and $1 \leq j_1 < \cdots < j_r \leq n$, and the matrix is in reduced row echelon form if it satisfies:

1. For $i = 1, \ldots, r$ and $j < j_i$, $E_{i,j} = 0$.
2. For $i = 1, \ldots, r$, $E_{i,j_i} = 1$.
3. For $i = 1, \ldots, r$ and $k < i$, $E_{k,j_i} = 0$.
4. For $i > r$ and any j, $E_{i,j} = 0$.

There is also the notion of ***row echelon form***, which are matrices that look like

$$\begin{pmatrix} \circledast & * & * & 0 & * & 0 & * & * \\ 0 & 0 & 0 & \circledast & * & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & \circledast & * & * \end{pmatrix}$$

where the entries marked by $\circledast$ are nonzero, and thus the pivots. A matrix is in row echelon form if it satisfies

1. For $i = 1, \ldots, r$ and $j < j_i$, $E_{i,j} = 0$ (all entries to the left of a pivot are 0).
2. For $i = 1, \ldots, r$, $E_{i,j_i} \neq 0$ (all pivots are nonzero).
3. For $i > r$ and any j, $E_{i,j} = 0$ (all rows below the pivot rows must consist only of zeros).

Now, we will discuss why matrices in reduced row echelon form are easier to work with. For instance, consider finding the null space of a linear map T. Suppose the matrix of T, denoted E, is in reduced row echelon form and has entries

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

To find $\text{Null}(T)$, we wish to find solutions to $Ex = \vec{0}$, which is equivalent to

$$\begin{pmatrix} 1 & 2 & 0 & 3 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \vec{0}.$$

We can write this as the system of equations

$$\begin{aligned} x_1 + 2x_2 + 3x_4 &= 0 \\ x_3 - x_4 &= 0 \\ x_5 &= 0. \end{aligned}$$

To solve these equations, we can let x_2 and x_4 take on arbitrary values, which would uniquely determine the values of x_1, x_3, and x_5. The variables x_2, x_4 are called ***free variables*** because they can be freely assigned to any value. The other variables x_1, x_3, x_5 are called ***pivot variables***, which correspond to the pivots of the matrix E. While the above example use homogeneous linear equations, the same analysis applies to inhomogeneous linear equations as well (which can be used to find $\text{Range}(T)$).

Note that the last equation of the above system of equations has the least number of variables (only x_5) while the first equation has the highest number of variables. In general, the goal of Gaussian elimination is to create a simplified system of linear equations and first solve the equations with the fewest variables, assigning values to free and pivot variables along the way, until the system is solved. We will dive into the specifics of Gaussian elimination in the next lecture.

10 Gaussian Elimination (continued)

10.1 Review

Last time, we introduced Gaussian elimination, which is an algorithm that starts with an $m \times n$ matrix A converts it into a matrix in reduced row echelon form. Recall that a matrix is in reduced row echelon form if it looks like:

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

where the leading 1s are called pivots.

10.2 Elementary Row Operations

As mentioned in the previous lecture, Gaussian elimination converts a matrix into reduced row echelon form by performing *elementary row operations*.

Definition 10.1 (elementary row operations)
The following are **elementary row operations**:

1. $M(i, c)$: Multiply the i^{th} row by a nonzero number $c \in \mathbb{F}$.
2. $A(i \xrightarrow{c} j)$: Add c times the ith row to the j^{th} row.
3. $S(i, j)$: Swap the i^{th} and j^{th} rows.

Example 10.2
Consider the matrix

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

The following are examples of elementary row operations on this matrix.

1. Multiplying the second row by $\frac{1}{2}$:

$$\begin{pmatrix} 0 & 1 & 2 \\ 1 & 2 & 5 \\ 2 & 1 & 0 \end{pmatrix} \xrightarrow{M(2;\frac{1}{2})} \begin{pmatrix} 0 & 1 & 2 \\ \frac{1}{2} & 1 & \frac{5}{2} \\ 2 & 1 & 0 \end{pmatrix}.$$

2. Adding -1 times the third row to the first row:

$$\begin{pmatrix} 0 & 1 & 2 \\ 1 & 2 & 5 \\ 2 & 1 & 0 \end{pmatrix} \xrightarrow{A(3 \xrightarrow{-1} 1)} \begin{pmatrix} -2 & 0 & 2 \\ 1 & 2 & 5 \\ 2 & 1 & 0 \end{pmatrix}.$$

3. Swapping the first and third rows:

$$\begin{pmatrix} 0 & 1 & 2 \\ 1 & 2 & 5 \\ 2 & 1 & 0 \end{pmatrix} \xrightarrow{S(1,3)} \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 5 \\ 0 & 1 & 2 \end{pmatrix}.$$

Gaussian elimination consists of performing a series of these three elementary row operations until we get our desired result.

10.3 Performing Gaussian Elimination

Finally, we will show how to actually perform Gaussian elimination. We will present the algorithm in two parts.

Algorithm 10.3 (Gaussian elimination)
Suppose A is a matrix. First, perform the following steps:

1. Find a leftmost entry in A (may not be unique) and let it be $a = A_{i,j}$. Apply $M(i; a^{-1})$ to make the pivot equal to 1 and $S(i, 1)$ to make it the first row. Let $j_1 = j$, so $(1, j_1)$ is the first pivot.
2. Look down from the first pivot. For each nonzero entry A_{i,j_1} for $i \geq 2$, apply $A(1 \xrightarrow{A_{i,j_1}^{-1}} i$ to make that entry equal to 0. Then, all entries below the first pivot are equal to 0.
3. Delete the first row and repeat the above steps.

The output of the above steps is a matrix in row echelon form with all pivots equal to 1.

Then, perform the following steps:

1. Start with the last pivot (r, j_r) and look above it. For each nonzero entry A_{s,j_r} for $s < r$, apply $A(r \xrightarrow{-A_{s,j_r}^{-1}} s$ to make that entry equal to 0. Then, all entries above the last pivot are equal to 0.
2. Repeat the above step for pivots $(r-1, j_{r-1}), \ldots, (1, j_1)$, in order.

The output of the entire algorithm is a matrix in reduced row echelon form.

As an in-depth example, we will perform the first part of Gaussian elimination on the matrix

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

We will take the first row to be our first pivot. Then, we perform the elementary row operations

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 0 \\ 2 & 2 & 1 \end{pmatrix} \xrightarrow{A(1 \xrightarrow{-2} 2)} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 2 & 2 & 1 \end{pmatrix} \xrightarrow{A(1 \xrightarrow{-2} 3)} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -2 & -5 \end{pmatrix}$$

to make all entries below the pivot equal to 0. Now, we ignore the first row and repeat the steps on the second and third rows. We take the second row as our pivot and apply

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -2 & -5 \end{pmatrix} \xrightarrow{M(2; -\frac{1}{3})} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & -2 & -5 \end{pmatrix}$$

to make the pivot entry equal to 1. Next, we apply

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & -2 & -5 \end{pmatrix} \xrightarrow{A(2 \xrightarrow{2} 3)} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & -1 \end{pmatrix}.$$

To make the entry below the pivot equal to 0. Finally, we apply

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & -1 \end{pmatrix} \xrightarrow{M(3; -1)} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$$

to make the last pivot equal to 1, and we are finished with the first part of Gaussian elimination. Note that the matrix is in row echelon form with all pivots equal to 1.

Now, we will perform the second part of Gaussian elimination, which is much more straightforward. Starting from the last pivot, we apply

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{A(3 \xrightarrow{-2} 2)} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{A(3 \xrightarrow{-3} 1)} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

to make all entries above the pivot equal to 0. Then, we look at the second pivot and apply

$$\begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{A(2 \xrightarrow{-2} 1)} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Now, the matrix is in reduced row echelon form and we have finished Gaussian elimination.[16]

Now that you understand the logic behind performing Gaussian elimination, consider the next example, which is how we will show performing Gaussian elimination from now on.

Example 10.4
Perform Gaussian elimination:

$$\begin{pmatrix} 0 & 0 & 0 & 2 \\ 0 & 1 & 2 & 4 \\ 0 & 3 & 6 & 9 \end{pmatrix} \xrightarrow{S(1,2)} \begin{pmatrix} 0 & 1 & 2 & 4 \\ 0 & 0 & 0 & 2 \\ 0 & 3 & 6 & 9 \end{pmatrix}$$
$$\xrightarrow{A(1 \xrightarrow{-3} 3)} \begin{pmatrix} 0 & 1 & 2 & 4 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & -3 \end{pmatrix} \xrightarrow{M(2;\frac{1}{2})} \begin{pmatrix} 0 & 1 & 2 & 4 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -3 \end{pmatrix}$$
$$\xrightarrow{A(2 \xrightarrow{3} 3)} \begin{pmatrix} 0 & 1 & 2 & 4 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow{A(2 \xrightarrow{-4} 1)} \begin{pmatrix} 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

We could have performed Gaussian elimination like this:

$$\begin{pmatrix} 0 & 0 & 0 & 2 \\ 0 & 1 & 2 & 4 \\ 0 & 3 & 6 & 9 \end{pmatrix} \xrightarrow{M(3;\frac{1}{3})} \begin{pmatrix} 0 & 0 & 0 & 2 \\ 0 & 1 & 2 & 4 \\ 0 & 1 & 2 & 3 \end{pmatrix}$$
$$\xrightarrow{S(1,3)} \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 0 & 2 \end{pmatrix} \xrightarrow{A(1 \xrightarrow{-1} 2)} \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$
$$\xrightarrow{A(2 \xrightarrow{-2} 3} \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow{A(2 \xrightarrow{-3} 1} \begin{pmatrix} 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

In Example 10.4, we had a choice to make at the start of Gaussian elimination; we could either take the second row or the third row as the first pivot. This begs the question: is the output of Gaussian elimination unique? In other words, is it possible to perform Gaussian elimination on the same matrix in two different ways and get two different outputs? Clearly, both our methods in Example 10.4 produced the same output. However, we leave the question on whether or not this is true in general as a thinking point for the reader.

10.4 Null Space and Range Using Reduced Row Echelon Form

After performing Gaussian elimination, we are left with a matrix in reduced row echelon form. In this section, we will show how to use matrices in reduced row echelon form to compute the null space and range.

Suppose A is an $m \times n$ matrix and R is the output of performing Gaussian elimination. Let R be in reduced row echelon form with pivot entries $(1, j_1), \ldots, (r, j_r)$. As a concrete example, say

$$R = \begin{pmatrix} 0 & 1 & * & 0 & 0 & * \\ 0 & 0 & 0 & 1 & 0 & * \\ 0 & 0 & 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We say that ***free indices*** are indices $1 \le i \le n$ that do not appear in $j_1, \ldots, j_r$.[17] In other words, free indices are the indices $1 \le i \le n$ such that the i^{th} column does not contain a pivot.

[16] Note that in the second part of Gaussian elimination, we don't actually need to write out all the elementary row operations. Instead, we can just set all entries above the pivots equal to 0, which would give the desired output.

[17] The notion of free indices is related to *free variables*, which we mentioned at the end of last lecture.

Let $S \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$ be a linear map defined by $R = \mathcal{M}(S)$. First, let's find Range(S). Let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{F}^n$. Then, $S(e_{j_i}), \ldots, S(e_{j_r})$ is a basis of Range(S). This is true because only the first r rows of R are nonzero, and the remaining rows consist only of zeroes. This implies that

$$\text{Range}(S) = \{(x_1, \ldots, x_r, 0, \ldots, 0) \in \mathbb{F}^m \mid x_1, \ldots, x_r \in \mathbb{F}\} \cong \mathbb{F}^r.$$

Additionally, let $f_1, \ldots, f_m$ be the standard basis of $\mathbb{F}^m$. By the matrix R, we see that

$$S(e_{j_i}) = i^{\text{th}} \text{ column of } R = f_i$$

for all $i = 1, \ldots, r$, which proves that $S(e_{j_i}), \ldots, S(e_{j_r})$ is a basis of Range(S).

Now, we will consider the original matrix A. The next result is key to why Gaussian elimination works.

Lemma 10.5
Suppose A is an $m \times n$ matrix and $T \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$ is a linear map defined by $A = \mathcal{M}(T)$. Let $e_1, \ldots, e_n$ be the standard basis of $\mathbb{F}^n$. Then, $T(e_{j_1}), \ldots, T(e_{j_r})$ is a basis of Range(T).

Proof. First, we will show that each elementary row operation is equivalent to left multiplying by an invertible matrix. We can see this by the fact that for any matrix A,

- $M(i;c)$ is equivalent to

$$A \xrightarrow{M(i;c)} \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & c & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} A.$$

- $A(i \xrightarrow{c} j)$ is equivalent to

$$A \xrightarrow{A(i \xrightarrow{c} j)} \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & c & & \\ & & & \ddots & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} A.$$

- $S(i,j)$ is equivalent to

$$A \xrightarrow{S(i,j)} \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} A.$$ [18]

Thus, Gaussian elimination is essentially starting with a matrix A and left multiplying it with a series of invertible matrices until we get a matrix R in reduced row echelon form. Additionally, note that the product of invertible matrices is also invertible, so multiplying by a series of invertible matrices can be rewritten as multiplying by a single, which we will denote as B.

[18] For more information, visit https://en.wikipedia.org/wiki/Elementary_matrix.

Now, we have Gaussian elimination starting with a matrix A and outputting $R = BA$, which implies B is a $m \times m$ matrix. Let $S \in \mathcal{L}(\mathbb{F}^m, \mathbb{F}^m)$ be a linear map defined by $B = \mathcal{M}(S)$. Then, the product BA is equivalent to the composition of linear maps ST :

$$\mathbb{F}^n \xrightarrow{T} \mathbb{F}^m \xrightarrow{S} \mathbb{F}^m.$$

We can relate Range(T) and Range(ST) by applying S,

$$\text{Range}(T) \xrightarrow{S} \text{Range}(ST)$$

which maps $v \in \text{Range}(T)$ to Sv. This map is well-defined because if $v \in \text{Range}(T)$, there exists $w \in \mathbb{F}^n$ such that $v = Tw$, so $Sv = STw \in \text{Range}(ST)$. Additionally, because B is invertible, it follows that S is an isomorphism.

Now, since $R = BA$ is in reduced row echelon form, we know from our previous discussion that $ST(e_{j_1}), \ldots, ST(e_{j_r})$ is a basis of Range(ST). Because S is an isomorphism, applying S^{-1} implies that $T(e_{j_1}), \ldots, T(e_{j_r})$ is a basis of Range(T), as desired. □

Example 10.6
Suppose A is the starting matrix in Example 10.4

$$A = \begin{pmatrix} 0 & 0 & 0 & 2 \\ 0 & 1 & 2 & 4 \\ 0 & 3 & 6 & 9 \end{pmatrix}$$

and T is the linear map with matrix A. After performing Gaussian elimination, we get the matrix

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

This matrix has pivots in the second and fourth columns which implies that $T(e_2), T(e_4)$ is a basis of Range(T) by Lemma 10.5, which is equivalent to the second and fourth columns of A. Therefore,

$$\left\{ \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \\ 9 \end{pmatrix} \right\}$$

is a basis of Range(T).

Now, we will prove a similar lemma with the null space.

Lemma 10.7
Suppose A is an $m \times n$ matrix and $T \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$ is a linear map defined by $A = \mathcal{M}(T)$. Let R be the matrix resulting from performing Gaussian elimination on A and $ST \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$ be the linear map defined by $R = \mathcal{M}(ST)$. Then, Null(T) = Null(ST).

Proof. Recall from the proof of Lemma 10.5 that S is an isomorphism. Thus, for any $v \in \mathbb{F}^n$, $Tv = 0$ if and only if $Sv = 0$, so Null(T) = Null(ST). □

Thus, finding Null(T) is equivalent to finding the null space of matrix R. However, finding the null space of R is not as simple as finding the range. Recall that ***free variables*** are variables x_j where j is a free index. As we discussed at the end of last lecture, the free variables can be assigned to any number, which uniquely determines the pivot variables $x_{j_1}, \ldots, x_{j_r}$. We will use the next example to explain how to find Null(T) from the pivot variables.

Example 10.8
Suppose we perform Gaussian elimination and get the matrix

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

To get the null space, this should be a homogeneous system of equations, which is equivalent to

$$\begin{aligned} x_1 + 2x_2 + 3x_4 &= 0 \\ x_3 - x_4 &= 0 \\ x_5 &= 0. \end{aligned}$$

Note that the free variables are x_2, x_4. We can express the pivot variables in terms of the free variables:

$$\begin{aligned} x_1 &= -2x_2 - 3x_4 \\ x_3 &= x_4 \\ x_5 &= 0. \end{aligned}$$

Thus, we can express the pivot variables only in terms of the free variables, which shows that after we assign the free variables, the pivot variables are uniquely determined. Additionally, any assignment of the free variables x_2, x_4 will correspond with a vector in $\mathrm{Null}(T)$.

To obtain a basis of $\mathrm{Null}(T)$, we plug in the assignments $(x_2, x_4) = (1, 0)$ and $(x_2, x_4) = (0, 1)$ and find the corresponding values of the pivot values. Computing, we find that

$$\left\{ \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -3 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \right\}$$

is a basis of $\mathrm{Null}(T)$.

We generalize the results of the above example with the following: suppose A is an arbitrary $m \times n$ matrix and $e_1, \ldots, e_n$ is the standard basis of $\mathbb{F}^n$. Let $j_1, \ldots, j_r$ be the pivot indices of A and let j be a free index of A. For each free index j, define

$$v_j = e_j - \sum_{i=1}^{r} R_{ij} e_{j_i}.$$

Then, the vectors v_j for all free indices j forms a basis of $\mathrm{Null}(T)$.

10.5 Solving Inhomogeneous Equations

Recall that an inhomogeneous system of equations is of the form $Ax = b$. In this section, we will show how to use Gaussian elimination to solve equations of this form.

Suppose A is a $m \times n$ matrix and $b \in \mathbb{F}^m$. First, form the ***augmented matrix***

$$(A|b),$$

which is the $m \times (n+1)$ matrix obtained by appending b to the end of A. Then, perform Gaussian elimination on A; at each step, apply the row operation to the augmented matrix $(A|b)$ instead of just to A. At the end, we are left with the augmented matrix

$$(R|d),$$

where R is in reduced row echelon form and $d \in \mathbb{F}^m$.

Example 10.9
Suppose A is the starting matrix in Example 10.4

$$A = \begin{pmatrix} 0 & 0 & 0 & 2 \\ 0 & 1 & 2 & 4 \\ 0 & 3 & 6 & 9 \end{pmatrix}$$

and we wish to solve

$$Ax = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}.$$

We form the augmented matrix $(A|b)$:

$$\left(\begin{array}{cccc|c} 0 & 0 & 0 & 2 & b_1 \\ 0 & 1 & 2 & 4 & b_2 \\ 0 & 3 & 6 & 9 & b_3 \end{array}\right).$$

Then, we perform Gaussian elimination on A, while applying the row operation to the entire augmented matrix $(A|b)$:

$$\left(\begin{array}{cccc|c} 0 & 0 & 0 & 2 & b_1 \\ 0 & 1 & 2 & 4 & b_2 \\ 0 & 3 & 6 & 9 & b_3 \end{array}\right) \xrightarrow{S(1,2)} \left(\begin{array}{cccc|c} 0 & 1 & 2 & 4 & b_2 \\ 0 & 0 & 0 & 2 & b_1 \\ 0 & 3 & 6 & 9 & b_3 \end{array}\right)$$
$$\xrightarrow{A(1 \xrightarrow{-3} 3)} \left(\begin{array}{cccc|c} 0 & 1 & 2 & 4 & b_2 \\ 0 & 0 & 0 & 2 & b_1 \\ 0 & 0 & 0 & -3 & -3b_2 + b_3 \end{array}\right) \xrightarrow{M(2;\frac{1}{2})} \left(\begin{array}{cccc|c} 0 & 1 & 2 & 4 & b_2 \\ 0 & 0 & 0 & 1 & \frac{1}{2}b_1 \\ 0 & 0 & 0 & -3 & -3b_2 + b_3 \end{array}\right)$$
$$\xrightarrow{A(2 \xrightarrow{3} 3)} \left(\begin{array}{cccc|c} 0 & 1 & 2 & 4 & b_2 \\ 0 & 0 & 0 & 1 & \frac{1}{2}b_1 \\ 0 & 0 & 0 & 0 & \frac{3}{2}b_1 - 3b_2 + b_3 \end{array}\right) \xrightarrow{A(2 \xrightarrow{-4} 1)} \left(\begin{array}{cccc|c} 0 & 1 & 2 & 0 & -2b_1 + b_2 \\ 0 & 0 & 0 & 1 & \frac{1}{2}b_1 \\ 0 & 0 & 0 & 0 & \frac{3}{2}b_1 - 3b_2 + b_3 \end{array}\right).$$

Now, we will show that solving $Ax = b$ is equivalent to solving $Rx = d$. As we showed earlier, there exists an invertible matrix B such that $R = BA$. Additionally, since we performed the same elementary row operations on b, it follows that $d = Bb$. Thus, left multiplying both sides of $Ax = b$ by B gives us $Rx = d$, and left multiplying both sides of $Rx = d$ by B^{-1} gives $Ax = b$, so the both equations are equivalent.

Thus, it remains to solve $Rx = d$. From earlier, we know that only the first r rows of R are nonzero, where r is the number of pivots. Let $e_1, \ldots, e_m$ be the standard basis of $\mathbb{F}^m$. Thus, a solution exists if and only if $d \in \text{span}(e_1, \ldots, e_r)$. If a solution exists, we can find a particular solution, denoted x_p, by setting all free variables equal to 0. Then, we can get the general solution by adding x_0 to the particular solution, where x_0 is in the null space of R. This is because

$$R(x_p + x_0) = Rx_p + Rx_0 = d + \vec{0} = d,$$

as desired.

Example 10.10

Continuing on Example 10.9, we wish to solve $Rx = d$, which is equivalent to

$$\begin{pmatrix} 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -2b_1 + b_2 \\ \frac{1}{2}b_1 \\ \frac{3}{2}b_1 - 3b_2 + b_3 \end{pmatrix}.$$

Since R has two pivots, it follows that a solution exists if and only if $d \in \operatorname{span}(e_1, e_2)$, which occurs when $\frac{3}{2}b_1 - 3b_2 + b_3 = 0$. Variables x_1, x_3 are free, so we can find a particular solution by setting them equal to 0, which gives

$$x_p = \begin{pmatrix} 0 \\ -2b_1 + b_2 \\ 0 \\ \frac{1}{2}b_1 \end{pmatrix}.$$

To find the general solution, we must calculate the null space of matrix A, which is left as an exercise to the reader.

11 Eigenvalues, Eigenvectors, and Invariant Subspaces

11.1 Review

Last time, we finished up our discussion of Gaussian elimination.

11.2 Eigenvalues and Eigenvectors

In this lecture, we will begin a new chapter that focuses on the properties of linear operators.

Definition 11.1 (operator)
A linear map from a vector space to itself is called an **operator**.

The set of all operators on V is denoted by $\mathcal{L}(V)$; in other words, $\mathcal{L}(V) = \mathcal{L}(V, V)$.

When expressing an operator in terms of a matrix, we almost always use the same basis for the domain and the codomain. Also, note that the matrix of an operator will be a square matrix.

Additionally, there is a special kind of matrices called ***diagonal matrices***, which are square matrices that have zeroes everywhere except possibly on the diagonal:

$$\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

Note that $\lambda_1, \dots, \lambda_n \in \mathbb{F}$ can either be zero or nonzero. Diagonal matrices are particularly easy to work with. For instance, multiplying two diagonal matrices consists of simply multiplying the corresponding entries:

$$\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \begin{pmatrix} \mu_1 & & 0 \\ & \ddots & \\ 0 & & \mu_n \end{pmatrix} = \begin{pmatrix} \lambda_1\mu_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n\mu_n \end{pmatrix}.$$

Now, suppose T is a linear operator. What does it mean for $\mathcal{M}(T)$ to be diagonal? Let $\lambda_1, \dots, \lambda_n$ be the diagonal entries of $\mathcal{M}(T)$. It follows that

$$\begin{aligned} T(v_1) &= \lambda_1 v_1 \\ &\vdots \\ T(v_n) &= \lambda_n v_n. \end{aligned}$$

Thus, when we apply T to any basis vector v_i, we get a multiple of v_i. This leads us to the notion of eigenvalues and eigenvectors.

Definition 11.2 (eigenvector, eigenvalue)
Suppose $T \in \mathcal{L}(V)$.

1. A vector $v \in V$ is called an **eigenvector** of T with eigenvalue $\lambda \in \mathbb{F}$ if $v \neq 0$ and $T(v) = \lambda v$.
2. A number $\lambda \in \mathbb{F}$ is called an **eigenvalue** if there is an eigenvector of T with eigenvalue λ. In other words, λ is an eigenvalue of there exists $v \in V$ such that $v \neq 0$ and $T(v) = \lambda v$.

One thing to note is that an eigenvalue can have more than one eigenvector corresponding to it, as seen in the following example.

Example 11.3
Suppose $T = \mathrm{id}_V$. Then, $T(v) = v$ for all $v \in V$. Thus, all vectors $v \in V$ such that $v \neq 0$ is an eigenvector of id_V with eigenvalue 1.

Do there exist any other eigenvalues of id_V? If $\lambda \in \mathbb{F}$ is an eigenvalue, then there exists some nonzero vector v such that $T(v) = \lambda v$. However, we know that $T(v) = v$, which implies that $\lambda v = v$. This is equivalent to $(1 - \lambda)v = \vec{0}$, which proves that $\lambda = 1$ because $v \neq 0$. Therefore, 1 is the only eigenvalue of id_V.

Now, let's consider the following question.

Guiding Question
When is 0 an eigenvalue of $T \in \mathcal{L}(V)$?

If 0 is an eigenvalue, it follows that there exists some vector $v \neq 0$ such that $T(v) = 0v = \vec{0}$. Thus, $v \in \text{Null}(T)$. Conversely, if $v \in \text{Null}(T)$ and $v \neq 0$, then $T(v) = \vec{0} = 0v$, so v is an eigenvector of T with eigenvalue 0. Therefore, 0 is an eigenvalue of T if and only if $\text{Null}(T)$ contains a nonzero vector (in other words, $\text{Null}(T) \neq \{\vec{0}\}$).

The next result expands the above reasoning to arbitrary eigenvalues.

Lemma 11.4
Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. Then, the following are equivalent:

1. λ is an eigenvalue of T;
2. $T - \lambda \text{id}_V$ is not injective;
3. $T - \lambda \text{id}_V$ is not surjective;
4. $T - \lambda \text{id}_V$ is not invertible;
5. $\text{rank}(T - \lambda \text{id}_V) < \dim V$.

Proof. First, we will show that (1) and (2) are equivalent. Suppose λ is an eigenvalue of T. Then, there exists a nonzero vector $v \in V$ such that $T(v) = \lambda v$. Thus, we compute

$$(T - \lambda \text{id}_V)(v) = T(v) - \lambda v = \vec{0},$$

so $v \in \text{Null}(T - \lambda \text{id}_V)$. It follows that $\text{Null}(T - \lambda \text{id}_V) \neq \{\vec{0}\}$, so $T - \lambda \text{id}_V$ is not injective. Now, suppose $T - \lambda \text{id}_V$ is not injective. Then, there exists some nonzero vector such that $v \in \text{Null}(T - \lambda \text{id}_V)$. This implies that $T(v) = \lambda v$, so λ is an eigenvalue of T.

For brevity, let $S = T - \lambda \text{id}_V \in \mathcal{L}(V)$. Next, we will show that (2) and (5) are equivalent. Note that S being not injective is equivalent to $\dim \text{Null}(S) > 0$. By Theorem 8.1,

$$\dim \text{Null}(S) + \text{rank}(S) = \dim V.$$

Thus, $\dim \text{Null}(S) > 0$ implies $\text{rank}(S) < \dim V$ and vice versa.

Next, we will show that (3) and (5) are equivalent. Note that S being not surjective is equivalent to $\text{Range}(S) \neq V$. It is clear that $\text{Range}(S) \neq V$ if and only if $\text{rank}(S) < \dim V$.

Finally, we will show that (2) and (5) are equivalent. If S is not injective, then S is not invertible. On the other hand, if S is not invertible, then either S is not injective or S is not surjective. If S is not injective, then we are done. Otherwise, S is not surjective. However, we have already shown that (2) and (3) are equivalent (because they are both equivalent to (5)), so S is not surjective if and only if S is not injective. Thus, S is not invertible implies S is not injective.

Therefore, all statements are equivalent. □

11.3 Finding Eigenvalues

Suppose T is a linear operator and $\mathcal{M}(T)$ is diagonal under $v_1, \ldots, v_n$. We now know that each v_i is an eigenvector of T with λ_1, where $\lambda_1, \ldots, \lambda_n$ are the diagonal entries of $\mathcal{M}(T)$. Thus, it is desirable to find eigenvectors of T, so that we *might* be able to find a basis under which $\mathcal{M}(T)$ is diagonal.

To start off simple, let's first try to find the eigenvalues and eigenvectors of an arbitrary 2×2 matrix. Suppose $T \in \mathcal{L}(V)$ and $\mathcal{M}(T) = A$ under the basis v_1, v_2, where A is defined as

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

Then, λ is an eigenvalue of T if and only if $\text{Null}(T - \lambda \text{id}_V) \neq \{\vec{0}\}$ by Lemma 11.4, which occurs if and only if the matrix

$$\mathcal{M}(T - \lambda \text{id}_V) = A - \lambda I = \begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix}$$

has nonzero null space. To determine the null space of the above matrix, we find the rank of the matrix by perform Gaussian elimination.

- If $a - \lambda \neq 0$,

$$\begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix} \xrightarrow{A(1 \xrightarrow{-\frac{c}{a-\lambda}} 2)} \begin{pmatrix} a-\lambda & b \\ 0 & \frac{(a-\lambda)(d-\lambda)-bc}{a-\lambda} \end{pmatrix}.$$

  Note that we did not finish Gaussian elimination because this simplification is enough for our purposes. Since $a - \lambda \neq 0$ by assumption, $\text{rank}(A - \lambda I) < 2$ if and only if the second row is entirely zeros, which occurs when $(a-\lambda)(d-\lambda) - bc = 0$.[19]

- If $a - \lambda = 0$ and $c \neq 0$,

$$\begin{pmatrix} 0 & b \\ c & d-\lambda \end{pmatrix} \xrightarrow{S(1,2)} \begin{pmatrix} c & d-\lambda \\ 0 & b \end{pmatrix}.$$

  Thus, $\text{rank}(A - \lambda I) < 2$ if and only if $b = 0$.

- If $a - \lambda = 0$ and $c = 0$, then $\text{rank}(A - \lambda I)$ must be less than 2 because it has an entire column of zeros.

In every case, we can determine that $\text{rank}(A - \lambda I) < 2$ if and only if $(a-\lambda)(d-\lambda) - bc = 0$, which can be rewritten as $\det(A - \lambda I) = 0$ (verify that this works for the second and third cases).

Consider the following example of a 2×2 matrix.

Example 11.5

Suppose V is a vector space over $\mathbb{F}$ and $T \in \mathcal{L}(V)$ such that $\mathcal{M}(T) = A$ for

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

We calculate that $\det(A - \lambda I) = \lambda^2 + 1$, so we wish to find solutions to the equation

$$\lambda^2 + 1 = 0$$

If $\mathbb{F} = \mathbb{C}$, then this equation has solutions $\lambda = \pm i$, so i and $-i$ are eigenvalues of T. On the other hand, if $\mathbb{F} = \mathbb{R}$, then this equation has no solutions, so T has no eigenvalues.

The above example shows that the eigenvalues of an operator T can depend on the field $\mathbb{F}$ that the vector space is over. Now, the next example will demonstrate how to find eigenvectors corresponding to a particular eigenvalue.

[19] You might recognize this expression as the determinant of $A - \lambda I$. The textbook for this course is notorious for not introducing determinants until very late in the book. For our course, knowing the formula for the determinant of a 2×2 matrix will be sufficient.

Example 11.6
Suppose T is defined as in Example 11.5 and $\mathbb{F} = \mathbb{C}$. If λ is an eigenvalue of T, then an eigenvector v corresponding to λ must be in the null space of the matrix

$$A - \lambda I = \begin{pmatrix} -\lambda & -1 \\ 1 & -\lambda \end{pmatrix}.$$

This is equivalent to solving the homogeneous equation

$$\begin{pmatrix} -\lambda & -1 \\ 1 & -\lambda \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

We know that T has eigenvalues i and $-i$. First, let's consider $\lambda = i$. The equation

$$\begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

must have at least one solution, since we proved in Example 11.5 that the rank of this matrix is less than 2. In particular, we can find that $(x_1, x_2) = (1, -i)$ is a solution, and thus an eigenvector of T with eigenvalue i. Similarly, we can find that $(x_1, x_2) = (1, i)$ is an eigenvector of T with eigenvalue $-i$.

Denote the eigenvectors as

$$v_1 = \begin{pmatrix} 1 \\ -i \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ i \end{pmatrix}.$$

Is it easy to see that v_1, v_2 is a basis of $\mathbb{C}^2$. Thus, based off our earlier discussion, $\mathcal{M}(T)$ with respect to basis v_1, v_2 should be the diagonal matrix

$$\mathcal{M}(T, (v_1, v_2)) = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix},$$

where the diagonal entries are the corresponding eigenvalues.

At this point in our discussion, we arrive at the following two questions.

Guiding Question
Suppose $T \in \mathcal{L}(V)$.

- How many eigenvalues does T have?
- Does there exist a basis of V consisting of eigenvectors of T?

For the first question, note that we have seen many examples of linear operators with different numbers of eigenvalues. For instance, the operator in Example 11.5 under $\mathbb{C}$ has $\dim V$ eigenvalues, while the same operator under $\mathbb{R}$ has 0 eigenvalues. Furthermore, we saw in Example 11.3 that id_V has 1 eigenvalue. In general, T can have as few as 0 eigenvalues and as many as $\dim V$ eigenvalues (which we will prove later).

For the second question, note that $\mathcal{M}(T)$ with respect to a basis of eigenvectors will be a diagonal matrix. In Example 11.6, we saw that there exists a basis of eigenvectors for that particular T. However, it does not necessarily exist for all operators T, as seen in the next example.

Example 11.7
Suppose $T \in \mathcal{L}(V)$ such that $\mathcal{M}(T) = A$ where

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

The determinant of $A - \lambda I$ is λ^2, so the only eigenvalue of T is 0.

Thus, if V had a basis v_1, v_2 of eigenvectors, it follows that both v_1, v_2 are eigenvectors with eigenvalue 0. However, $Tv_1 = Tv_2 = \vec{0}$ implies that $T = 0$, which is a contradiction. Therefore, there does not exist a basis of eigenvectors of T.

Now, we will prove the following result, which will limit the number of eigenvalues an operator can have.

Lemma 11.8
Suppose $T \in \mathcal{L}(V)$ and $v_1, \ldots, v_r$ are eigenvectors of T with distinct eigenvalues $\lambda_1, \ldots, \lambda_r$. Then, $v_1, \ldots, v_r$ are linearly independent.

Proof. Let $a_1, \ldots, a_r$ be scalars such that

$$a_1 v_1 + \cdots + a_r v_r = \vec{0}.$$

If we apply T to both sides, the right-hand side becomes $T(\vec{0}) = \vec{0}$. The left-hand side becomes

$$T(a_1 v_1 + \cdots + a_r v_r) = a_1 T(v_1) + \cdots + a_r T(v_r) = a_1 \lambda_1 v_1 + \cdots + a_r \lambda_r v_r.$$

Thus, we get the equation $a_1 \lambda_1 v_1 + \cdots + a_r \lambda_r v_r = \vec{0}$. However, we can apply T to both sides of this equation to get $a_1 \lambda_1^2 v_1 + \cdots + a_r \lambda_r^2 v_r = \vec{0}$. In general, we can apply T j times to get the equation

$$a_1 \lambda_1^j v_1 + \cdots + a_r \lambda_r^j v_r = \vec{0}.$$

Thus, we have an infinite number of linear independence relations. For the sake of contradiction, assume there exists a relation where $a_1, \ldots, a_r$ are not all zero and suppose $a_1 v_1 + \cdots + a_r v_r = \vec{0}$ is the relation with the least number of nonzero a_i's. We can apply T to both sides to get the relation $a_1 \lambda_1 v_1 + \cdots + a_r \lambda_r v_r = \vec{0}$. Without loss of generality, suppose $a_1 = 0$. We multiply our original relation by λ_1 to get

$$a_1 \lambda_1 v_1 + \cdots + a_r \lambda_l v_r = \vec{0}.$$

We subtract our second relation to get

$$a_1(\lambda_1 - \lambda_1) v_1 + \cdots + a_r(\lambda_1 - \lambda_r) v_r = \vec{0}.$$

Now, we will show that the coefficients of this relation $a_1(\lambda_1 - \lambda_1), \cdots, a_r(\lambda_1 - \lambda_r)$ has fewer nonzero coefficients than $a_1, \ldots, a_r$. First, note that $a_i(\lambda_1 - \lambda_i) = 0$ if and only if $a_i = 0$ for $i = 2, \ldots, r$. However, $a_1(\lambda_1 - \lambda_1) = 0$ while $a_i \neq 0$, so this new relation has one less nonzero coefficient. This is a contradiction, unless $a_2(\lambda_1 - \lambda_2) = \cdots = a_r(\lambda_1 - \lambda_r) = 0$. This implies that $a_2 = \cdots = a_r = 0$. Plugging this back into $a_1 v_1 + \cdots + a_r v_r = \vec{0}$, it follows that $a_1 = 0$, which is a contradiction. Therefore, $v_1, \ldots, v_r$ are linearly independent. □

A consequence of this lemma is that an operator $T \in \mathcal{L}(V)$ can have at most $\dim V$ distinct eigenvalues. To prove this, let $\lambda_1, \ldots, \lambda_r$ be all eigenvalues of T. For each eigenvalue λ_i, there exists at least one eigenvector v_i corresponding to λ_i. By Lemma 11.8, $v_1, \ldots, v_r$ is linearly independent. A list of linearly independent vectors has length at most $\dim V$, so $r \leq \dim V$.

11.4 Invariant Subspaces

We have seen how applying an operator T to an eigenvector sends it to a multiple of itself. What if we instead considered a subspace, which when applying an operator T, the subspace gets sent to itself? This motivates the notion of invariant subspaces.

Definition 11.9 (invariant subspace)
Suppose $T \in \mathcal{L}(V)$. A subspace U of V is called **T-invariant** if $T(u) \in U$ for all $u \in U$.

A simple example of invariant subspaces is as follows: let $v \in V$ and let $\mathbb{F} \cdot v$ denote the subspace of V consisting of multiples of v. Then, $\mathbb{F} \cdot v$ is T-invariant if and only if v is an eigenvector of T.

Now, suppose $T \in \mathcal{L}(V)$ and U is a T-invariant subspace of V. Choose a basis $u_1, \ldots, u_m$ of U and extend it to a basis $u_1, \ldots, u_m, u_{m+1}, \ldots, u_n$ of V. Note that when we express

$$T(u_1) = a_1 u_1 + \cdots + a_m u_m + a_{m+1} u_{m+1} + \cdots + a_n u_n,$$

the coefficients $a_{m+1}, \ldots, a_n$ will all be equal to 0 because $T(u_1) \in U$ by the definition of invariant subspaces. This reasoning holds for $T(u_1), \ldots, T(u_m)$. Thus, $\mathcal{M}(T)$ with respect to this basis would look like

$$\mathcal{M}(T) = \left(\begin{array}{ccc|ccc} * & \cdots & * & * & \cdots & * \\ \vdots & & \vdots & \vdots & & \vdots \\ * & \cdots & * & * & \cdots & * \\ \hline 0 & \cdots & 0 & * & \cdots & * \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & * & \cdots & * \end{array}\right)$$

This is called a ***block upper-triangular*** matrix, because the blocks of the matrix look upper-triangular. The upper-left block is the matrix of the ***restriction operator*** $T|_U \in \mathcal{L}(U)$. Additionally, the lower-right block is the matrix of another linear operator related to T which has to do with the ***quotient space***. Both of these will be explored further in the next lecture.

12 Invariant Subspaces (continued) and Upper-triangular Matrices

12.1 Review

Last time, we introduced the notion of eigenvectors and eigenvalues and we ended with a definition of invariant subspaces. Recall that if T is a linear operator acting on some vector space V, then subspace $U \subset V$ is T-invariant if $T(u) \in U$ for all $u \in U$ (Definition 11.9).

Furthermore, we showed that if we have a basis $u_1, \ldots, u_m$ of U and we extend it to a basis $u_1, \ldots, u_m, w_1, \ldots, w_k$ of V, then the matrix of T under this extended basis looks like the block upper-triangular matrix

$$\mathcal{M}(T) = \left(\begin{array}{c|c} * & * \\ \hline 0 & * \end{array}\right). \tag{3}$$

Recall that $\dim U = m$ and $\dim V = m+k$, which follows from how we defined the bases of U and V. Furthermore, we showed at the end of last lecture that the upper-left and bottom-right blocks are both square matrices with dimensions $m \times m$ and $k \times k$, respectively.

Guiding Question

Do the upper-left and bottom-right blocks represent some operator? If so, what are these operators relation to T?

12.2 Restriction Operator

First, we will consider the upper-left block of $\mathcal{M}(T)$.

Definition 12.1 (restriction operator)
Suppose $T \in \mathcal{L}(V)$ and U is a T-invariant subspace of V. Then, the **restriction operator** $T|_U \in \mathcal{L}(U)$ is defined by

$$T|_U(u) = T(u)$$

for all $u \in U$.

Essentially, $T|_U$ is identical to T except the domain and codomain of $T|_U$ is restricted to the subspace U instead of all of V.

Thus, the upper-left block is exactly the matrix of $\mathcal{M}(T|_U)$ under the basis $u_1, \ldots, u_m$. So, in terms of matrices, restricting your operator to a T-invariant subspace $U \subset V$ is equivalent to concentrating on the upper-left corner of $\mathcal{M}(T)$. We can now express $\mathcal{M}(T)$ as the block matrix

$$\mathcal{M}(T) = \left(\begin{array}{c|c} \mathcal{M}(T|_U) & * \\ \hline 0 & * \end{array}\right). \tag{4}$$

12.3 Quotient Space

Now, consider the lower-right block of $\mathcal{M}(T)$. Note that the lower-right block is a $k \times k$ square matrix, where $k = \dim V - \dim U$. So, we first have to define a vector space with dimension equal to $\dim V - \dim U$.

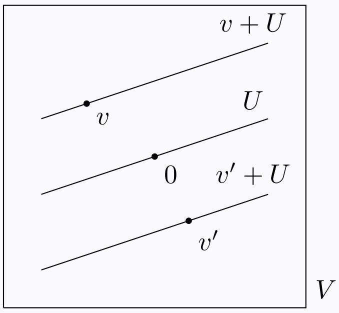
Definition 12.2 (quotient space)
The **quotient space** V/U [a] is the set of equivalence classes in V where two vectors v_1 and v_2 are equivalent if and only if $v_1 - v_2 \in U$.

[a] V/U is also sometimes denoted as $V \bmod U$.

This is analogous to decomposing the integers into odd and even integers: two integers are equivalent if and only if their difference is an even number. But, in the abstract case, instead of integers we are talking about vectors, which are equivalent if and only if their different lies in some smaller subspace.

Thus, if two vectors in V are equivalent according the above definition, we view them as the same element in V/U. Elements in V/U are denoted as $v+U$, where $v+U$ represents the subset of V containing the vectors $v+u$ for all $u \in U$. Essentially, $v+U$ can be thought of as collections of equivalent vectors, where v is simply a representative of the equivalence class.

Example 12.3
Consider a geometric example, where V is the 2D-plane and U is a line.



We can start at v and add anything in U, which forms a parallel line. Namely, this parallel line is $v+U$. If we pick a different v', we get a different parallel line representing $v'+U$. Thus, V/U is the set of lines in V parallel to U.

Now, we can define addition and scalar multiplication on V/U.

Definition 12.4 (addition and scalar multiplication on V/U)
Addition and scalar multiplication on V/U is defined by

- addition : $(v_1+U)+(v_2+U)=(v_1+v_2)+U$
- scalar multiplication : $c(v+U)=(cv)+U$

However, an issue with this definition is that the representation of any $v+U$ is not unique: if $v-v' \in U$, then $v+U=v'+U$. So, we must show that the definitions above are well-defined.

To show that addition is well-defined, we must show that if $v_1+U=v_1'+U$ and $v_2+U=v_2'+U$, then $(v_1+v_2)+U=(v_1'+v_2')+U$. Since $v_1+U=v_1'+U$ and $v_2+U=v_2'+U$, it follows that $v_1'-v_1 \in U$ and $v_2'-v_2 \in U$. But, U is closed under addition because U is a subspace, so it follows that $(v_1'-v_1)+(v_2'-v_2)=(v_1'+v_2')-(v_1+v_2) \in U$, which implies that $(v_1+v_2)+U=(v_1'+v_2')+U$.

Similarly, to show that scalar multiplication are well-defined, we must show that if $v+U=v'+U$, then $(cv)+U=(cv')+U$. Since $v'-v \in U$ and U is closed under scalar multiplication (recall U is a subspace), it follows that $c(v'-v)=cv'-cv \in U$, so $(cv)+U=(cv')+U$.

Now that addition and scalar multiplication on V/U are well-defined, we can show that V/U is a vector space. It is easy to verify that V/U under these operations satisfies the necessary axioms for a vector space.[20]

Now, we will focus our attention on the dimension of V/U.

Theorem 12.5 (Dimension of a quotient space)
Suppose V is a vector space and U is a subspace of V. Then

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

Proof. Construct a basis $u_1,\ldots,u_m$ of U and extend it to a basis $u_1,\ldots,u_m,w_1,\ldots,w_k$ of V. We will show that

[20] Note that the additive identity in V/U is $0+U$ and the additive inverse is $(-v)+U$.

$w_1 + U, w_2 + U, \dots, w_k + U$ is a basis of V/U.

First, we will show that $w_1 + U, w_2 + U, \dots, w_k + U$ span V/U. Take any $v + U \in V/U$. We can express $v = \sum_{i=1}^{n} a_i u_i + \sum_{j=1}^{k} b_j w_j$. Then, $v - (b_1 w_1 + \cdots + b_k w_k) = a_1 u_1 + \cdots a_m u_m \in U$, since $u_1, \dots, u_m$ form a basis of U. Thus,

$$v + U = b_1(w_1 + U) + b_2(w_2 + U) + \cdots + b_k(w_k + U),$$

so $w_1 + U, w_2 + U, \dots, w_k + U$ span V/U.

Next, we will check linear independence. Suppose we have the linear dependence relation $\sum_{j=1}^{k} b_j(w_j + U) = \vec{0}_{V/U}$, where $\vec{0}_{V/U}$ represents the zero vector in V/U. This implies that $\sum_{j=1}^{k} b_j w_j - \vec{0}_V \in U$, where $\vec{0}_V$ represents the zero vector in V, so $\sum_{j=1}^{k} b_j w_j \in U$. Thus, there exist constants a_i such that $\sum_{j=1}^{k} b_j w_j = \sum_{i=1}^{m} a_i u_i$. However, since $u_1, \dots, u_m, w_1, \dots, w_k$ is a basis, it follows that $a_i = b_j = 0$ for all i, j.

Therefore, $w_1 + U, w_2 + U, \dots, w_k + U$ is a basis of V/U, so $\dim V/U = k$. We know that $\dim V - \dim U = (m + k) - m = k$, so $\dim V/U = \dim V - \dim U$. □

Example 12.6

- $V/\{0\} = V$
- $V/V = \{0\}$

Remark. *Do not confuse V/U with the complement of U in V (denoted $V \setminus U$). For instance, $V/\{0\}$ is the entire vector space V, while $V \setminus \{0\}$ is V with the zero vector removed.*

Now, we will examine a linear map from V to V/U.

Definition 12.7 (quotient map)
Suppose U is a subspace of V. Then, the **quotient map** $\pi : V \to V/U$ is the linear map defined by

$$\pi(v) = v + U$$

for all $v \in V$.

Guiding Question
What are $\mathrm{Null}(\pi)$ and $\mathrm{Range}(\pi)$?

Answer. *For $\mathrm{Null}\,\pi$, is it clear that $\pi(v) = 0 + U$ if and only if $v \in U$, so $\mathrm{Null}\,\pi = U$.*

For $\mathrm{Range}\,\pi$, any element in V/U can be expressed as $v + U$ for some $v \in V$, so π is surjective. Thus, $\mathrm{Range}\,\pi = V/U$.

Remark. *The quotient map π also gives us an simpler way to prove Theorem 12.5. By the Fundamental Theorem of Linear Maps,*

$$\dim V = \dim \mathrm{Null}\,\pi + \dim \mathrm{Range}\,\pi = \dim U + \dim V/U,$$

so $\dim V/U = \dim V - \dim U$.[21]

Now, consider the composition of linear maps

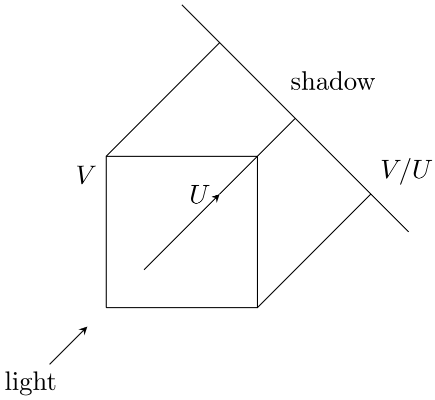
$$U \xrightarrow{\mathrm{id}} V \xrightarrow{\pi} V/U.$$

Guiding Question
What is the linear map from $U \to V/U$ given by the composition of id and π?

[21] This proof was not covered in lecture, although it is given in Axler.

Answer. *For* $u \in U$, *we know that* $\text{id}(u) = u \in U \subset V$. *Since* $\text{Null}(\pi)$, *it follows that* $\pi(\text{id}(u)) = 0$. *So, the composition of the two maps is the zero map.*

This structure helps illuminate the relation of U and V/U to V. While U is a part of V, the elements of V/U can be described as "shadows" of elements of V. For instance, consider the following diagram:



Imagine you have some directional light source, which shines on some arbitrary object. Then, the light points in some direction and collapses every point in that direction to a single point in the shadow. This example is analogous to quotient spaces, where V is analogous to the arbitrary object, U is analogous to the direction of the light, and V/U is analogous to the shadow of the object.

12.4 Quotient Operator

Finally, we are ready to return to our Guiding Question at the start of the lecture: what operator does the bottom-right block of $\mathcal{M}(T)$ represent?

Definition 12.8 (quotient operator)
Suppose $T \in \mathcal{L}(V)$ and U is a T-invariant subspace of V. Then, the **quotient operator** $T/U \in \mathcal{L}(V/U)$ is defined by

$$(T/U)(v + U) = Tv + U$$

for all $v \in V$.

First, we must show that the quotient operator is well-defined: that is, we need to verify that if $v + U = v' + U$, then $Tv + U = Tv' + U$. This is equivalent to showing that $Tv' - Tv = T(v' - v) \in U$, which is true because $v' - v \in U$ and U is a T-invariant subspace. So, the quotient operator is well-defined.

Now, recall that $u_1, \ldots, u_m$ is a basis of U and $u_1, \ldots, u_m, w_1, \ldots, w_k$ is a basis of V. In our proof of Theorem 12.5, we showed that $w_1 + U, \ldots, w_k + U$ is a basis of V/U.

Consider $\mathcal{M}(T/U)$ with respect to the basis $w_1 + U, \ldots, w_k + U$. This is exactly the lower-right block of $\mathcal{M}(T)$.Thus, we now have the completed picture of $\mathcal{M}(T)$, which is

$$\mathcal{M}(T) = \left(\begin{array}{c|c} \mathcal{M}(T|_U) & * \\ \hline 0 & \mathcal{M}(T/U) \end{array}\right). \tag{5}$$

12.5 Upper-Triangular Matrices

We will now shift our discussion to the topic of upper-triangular matrices.

Definition 12.9 (Upper-triangular matrix)
A matrix is called **upper-triangular** if all entries below the diagonal are 0.

Note that that the notion of upper-triangular matrices only applies to square matrices.

Example 12.10
The 5×5 matrix

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

is an upper-triangular matrix. The 0's below the diagonal are often shorthanded by a single big 0, so the matrix above is often written as

$$\begin{pmatrix} * & * & * & * & * \\ & * & * & * & * \\ & & * & * & * \\ & \text{\Large 0} & & * & * \\ & & & & * \end{pmatrix}.$$

Now, we will discuss the usefulness of upper-triangular matrices in finding eigenvalues.

Proposition 12.11
Suppose $T \in \mathcal{L}(V)$ and $\mathcal{M}(T)$ is upper-triangular under some basis of V. Then, the eigenvalues of T are the diagonal entries of $\mathcal{M}(T)$.

Example 12.12
Consider the matrix $A = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$. Then,

$$A \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

so the eigenvalues of A are $\lambda = 1, 2$.

So, if a basis of T exists such that $\mathcal{M}(T)$ with respect to that basis is upper-triangular, then the eigenvalues of T can be easily identified.

Let $\text{Eigen}(T)$ represent the set of eigenvalues of T. To prove Proposition 12.11, we will first need the following lemma.

Lemma 12.13
Suppose $T \in \mathcal{L}(V)$ and U is a T-invariant subspace of V. Then,

$$\text{Eigen}(T) = \text{Eigen}(T|_U) \cup \text{Eigen}(T/U).$$

Example 12.14
Consider the same matrix A as in Example 12.12. Then, $\mathcal{M}(T|_U) = \begin{pmatrix} 1 \end{pmatrix}$ and $\mathcal{M}(T/U) = \begin{pmatrix} 2 \end{pmatrix}$, which have eigenvalues $\lambda = 1$ and $\lambda = 2$, respectively.

Proof of Lemma 12.13. Let $\lambda \in \mathbb{F}$. Recall that $\lambda \in \text{Eigen}(T)$ if and only if $T - \lambda \text{id}_V$[22] is not invertible. Furthermore, $\lambda \in \text{Eigen}(T|_U) \cup \text{Eigen}(T/U)$ if and only if either $T|_U - \lambda \text{id}_V$ is not invertible or $T/U - \lambda \text{id}_V$ is not invertible. Now, let $S = T - \lambda \text{id}_V \in \mathcal{L}(V)$. It is easy to verify that U is S-invariant.Thus, $S|_U = T|_U - \lambda \text{id}_V$ and $S/U = T/U - \lambda \text{id}_V$.

It remains to show that S is not invertible if and only if either $S|_U - \lambda \text{id}_V$ or $S/U - \lambda \text{id}_V$ is not invertible. This is equivalent to showing that S is invertible if and only if both $S|_U$ and S/U are invertible.

[22] Recall that id_V is the identity operator in V.

First, suppose that S is invertible. Then, there exists some $R \in \mathcal{L}(V)$ such that $SR = RS = \mathrm{id}_V$. Since S is invertible and U is S-invariant, for any $v \in U$, there exists some $v' \in U$ such that $Sv' = v$. Then,

$$Rv = R(Sv') = \mathrm{id}_V(v') = v',$$

so U is R-invariant. Now, we will show that $R|_U$ is the inverse of $S|_U$ and R/U is the inverse of S/U. For any $v \in U$,

$$S|_U R|_U v = SRv = v,$$

so $R|_U$ is the inverse of $S|_U$. Furthermore, for any $v \in V$,

$$(S/U)(R/U)(v + U) = (S/U)(Rv + U) = SRv + U = v + U,$$

so R/U is the inverse of S/U. Therefore, both $S|_U$ and S/U are invertible.

Now, suppose that both $S|_U$ and S/U are invertible. Let $u_1, \dots, u_m$ be a basis of U and extend it to a basis $u_1, \dots, u_m, w_1, \dots, w_k$ of V. We will show that $S(u_1), \dots, S(u_m), S(w_1), \dots, S(w_k)$ is linearly independent.

Suppose $\sum_{i=1}^m a_i S(u_i) + \sum_{j=1}^k b_j S(w_j) = 0$. We know that $\sum_{i=1}^m a_i S(u_i) \in U$, so $\sum_{j=1}^k b_j S(w_j) = -\sum_{i=1}^m a_i S(u_i) \in U$. This implies that

$$\sum_{j=1}^k b_j (S/U)(w_j + U) = 0 + U.$$

Since (S/U) is invertible, it maps a basis of V/U to another basis. Thus, $(S/U)(w_1 + U), \dots, (S/U)(w_k + U)$ is a basis of V/U, so $b_j = 0$ for all j. We now have

$$\sum_{i=1}^m a_i S(u_i) = 0.$$

Similarly, since $u_i, \dots, u_m$ is a basis of U and $S|_U$ is invertible, it follows that $S(u_1), \dots, S(u_m)$ is a basis of U. Therefore, $a_i = 0$ for all i, so $S(u_1), \dots, S(u_m), S(w_1), \dots, S(w_k)$ is linearly independent.

Now, since $S(u_1), \dots, S(u_m), S(w_1), \dots, S(w_k)$ is linearly independent and $\dim V = m + k$, it must be a basis of V. Therefore, S maps a basis of V to another basis of V, so S is an isomorphism and thus is invertible. □

Finally, we are ready to prove Proposition 12.11.

Proof of Proposition 12.11. Consider the decomposition of $\mathcal{M}(T)$ into the blocks

$$\mathcal{M}(T) = \left(\begin{array}{c|ccc} \lambda_1 & * & \cdots & * \\ \hline 0 & \lambda_2 & & * \\ \vdots & & \ddots & \\ 0 & 0 & & \lambda_n \end{array}\right).$$

It is clear that $Tv_1 = \lambda_1 v_1$, so $\lambda_1 \in \mathrm{Eigen}(T)$. Let $U = \mathrm{span}(v_1)$, which is clearly a T-invariant subspace. Denote the lower-right block of $\mathcal{M}(T)$ as $\mathcal{M}(T)'$. By Equation 5,

$$T|_U = \begin{pmatrix} \lambda_1 \end{pmatrix} \text{ and } T/U = \mathcal{M}(T)'.$$

Then, by Lemma 12.13,

$$\begin{aligned} \mathrm{Eigen}(T) &= \mathrm{Eigen}(T|_U) \cup \mathrm{Eigen}(T/U) \\ &= \{\lambda_1\} \cup \mathrm{Eigen}(\mathcal{M}(T)'). \end{aligned}$$

Since $\mathcal{M}(T)'$ is also upper-triangular, we can continue in this fashion to show that $\lambda_2, \lambda_3, \dots, \lambda_n$ are eigenvalues of T. Therefore, the eigenvalues of T are the diagonal entries of $\mathcal{M}(T)$. □

13 Existence of Eigenvalues and Diagonal Matrices

13.1 Review

Last time, we ended with the statement that if A is an upper-triangular matrix, then the eigenvalues of A are the diagonal entries. However, we have not yet discussed the case where a linear operator T is not guaranteed to have an eigenvalue. The following theorem will show when we are guaranteed to have an eigenvalue.

13.2 Existence of Eigenvalues (Complex Vector Spaces)

Theorem 13.1 (Operators on complex vector spaces have an eigenvalue)
Suppose V is a finite-dimensional vector space over $\mathbb{C}$ and T is a linear operator in $\mathcal{L}(V)$. Then, T has an eigenvalue.

Note that this theorem doesn't tell you how to find an eigenvalue, but only asserts that there one exists.

Before we begin the proof, we must first discuss some properties of polynomials. Let $p(z)$ be a polynomial in $\mathcal{P}(\mathbb{C})$. We know how to evaluate a polynomial at some number a, which results in some complex number $p(a) \in \mathbb{C}$. But, we can also evaluate polynomials on matrices. For instance, if A is a square matrix, then $p(A)$ is also a square matrix.

Similarly, for any $T \in \mathcal{L}(V)$, we can evaluate $p(T)$, which will also be in $\mathcal{L}(V)$.

Example 13.2
Let $p(z) = \frac{1}{2}z^2 + 1$. Then,

$$p(A) = \frac{1}{2}A^2 + I,$$
$$p(T) = \frac{1}{2}T^2 + \mathrm{id}_V.$$

Now, we will break up the proof of Theorem 13.1 into multiple steps:

1. There exists a nonzero polynomial $p(z) \in \mathcal{P}(\mathbb{C})$ such that $p(T) = 0$.
2. If there exists a nonzero polynomial $p(z) \in \mathcal{P}(\mathbb{C})$ such that $p(T) = 0$, then T has an eigenvalue.

Proof of (1). Before proving the general statement, we will look at a few special cases.

Example 13.3
Consider the special case where $\dim V = 1$. It follows that every operator in $\mathcal{L}(V)$ is scalar multiplication by some number $\lambda \in \mathbb{C}$. Thus, in this case, T is equivalent to some number λ.

Thus, our problem has reduced to the trivial problem finding a polynomial $p(z)$ such that $p(\lambda) = 0$. One such polynomial is $p(z) = z - \lambda$.

Example 13.4

Consider the case where $\dim V = 2$. Then, we can convert T into its matrix form

$$\mathcal{M}(T) = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Let $p(z) = a_0 + a_1 z + \cdots + a_n z^n$. We wish to find $a_0, a_1, \ldots, a_n$ such that

$$a_0 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + a_1 \begin{pmatrix} a & b \\ c & d \end{pmatrix} + \cdots + a_n \begin{pmatrix} a & b \\ c & d \end{pmatrix}^n = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Now, consider the polynomial $p(z) = z^2 - (a+d)z + (ad - bc)$ [a]. The simple calculation to show that $p(z)$ is a solution to the 2-dimensional case is left to the reader.

[a] The motivation to behind this solution will be covered near the end of the course.

Now, we will prove the general case. Consider the following list of an infinite number of operators

$$\mathrm{id}_V, T, T^2, T^3, \ldots \in \mathcal{L}(V).$$

Since V is a finite-dimensional vector space, let $\dim V = n$. It follows that $\dim \mathcal{L}(V) = n^2$, so $\mathcal{L}(V)$ is finite-dimensional. Thus, for $N > n^2$, it must be that the list of operators $\mathrm{id}_V, T, T^2, \ldots, T^N$ is linearly dependent. This implies that there exists a linear combination

$$a_0 \mathrm{id}_V + a_1 T + a_2 T^2 + \cdots + a_N T^N = 0$$

for some $a_0, a_1, \ldots, a_N \in \mathbb{C}$ which are not all zero.

Let $p(z) = a_0 + a_1 z + \cdots + a_n z^N$. It is clear that $p(z)$ is nonzero and $p(T) = 0$. □

Remark. *Note that tells you a very ine ective way of finding $p(z)$, since it requires you to calculate $T, T^2, \ldots, T^N$ and somehow find scalars $a_0, a_1, \ldots, a_N$. This proof is mainly only useful for proving the existence of $p(z)$, not for finding $p(z)$.*

Before we begin the proof of (2), recall the Fundamental Theorem of Algebra states that every non-constant polynomial $p(z) \in \mathcal{P}(\mathbb{C})$ has a root in $\mathbb{C}$. Let $p_n(z)$ be a polynomial of degree n. Then, it follows that $p_n(z)$ has some root λ_1, so

$$p_n(z) = (z - \lambda_1) p_{n-1}(z).$$

It is clear that this procedure can be repeated on $p_{n-1}(z), p_{n-2}(z), \ldots, p_1(z)$ to get

$$p_n(z) = c(z - \lambda_1)(z - \lambda_2) \cdots (z - \lambda_n),$$

where $c \in \mathbb{C}$. Thus, another way of stating the Fundamental Theorem of Algebra is that any complex polynomial $p(z)$ can be factored into linear factors (this factorization is also unique).

Proof of (2). Consider some nonzero $p(z) \in \mathcal{P}(\mathbb{C})$ such that $p(T) = 0$. It is clear that $p(z)$ is a non-constant polynomial because $p(z)$ is nonzero and $p(T) = 0$. Now, we can apply the Fundamental Theorem of Algebra to conclude that

$$p(z) = c(z - \lambda_1)(z - \lambda_2) \cdots (z - \lambda_n)$$

for some nonzero $c \in \mathbb{C}$. Then,

$$p(T) = c(T - \lambda_i \mathrm{id}_V)(T - \lambda_2 \mathrm{id}_V) \cdots (T - \lambda_N \mathrm{id}_V) = 0.$$

Since $c \neq 0$, it follows that $(T - \lambda_i \mathrm{id}_V)(T - \lambda_2 \mathrm{id}_V) \cdots (T - \lambda_N \mathrm{id}_V) = 0$. Denote $S_i = T - \lambda_i \mathrm{id}_V \in \mathcal{L}(V)$ for $i = 1, 2, \ldots, N$. It follows that

$$S_1 S_2 \cdots S_N = 0.$$

If all $S_1, S_2, \ldots, S_N$ were invertible, then their product $S_1 S_2 \cdots S_N = 0$ would also be invertible, which is a contradiction. Thus, say S_i is not invertible. It follows that $S_i = T - \lambda_i \mathrm{id}_V$ is not invertible, so λ_i is an eigenvalue of T. □

Remark. *If you can find some $p(z)$ such that $p(T) = 0$, then an eigenvalue of T must be one of the roots of $p(z)$.*

Corollary 13.5
Suppose $T \in \mathcal{L}(V)$ where V is a vector space over $\mathbb{C}$. Then, there exists a basis of V such that $\mathcal{M}(T)$ is upper triangular.

Proof. We will prove this by induction on the dimension of V.

First, consider the base case $\dim V = 1$. Then, as shown in Example 13.3, T is equivalent to scalar multiplication by some $\lambda \in \mathbb{C}$. It follows that the matrix form of T is the 1×1 matrix $\mathcal{M}(T) = \begin{pmatrix}\lambda\end{pmatrix}$, which is upper-triangular.

Now, let $\dim V = n > 1$. Assume that the statement has been proved for all $\dim V < n$. By Theorem 13.1, T has an eigenvalue λ_1. This implies that T has an eigenvector v_1 such that $Tv_1 = \lambda v_1$. Let $U = \operatorname{span}(v_1) \subset V$. Since v_1 is an eigenvector, it follows that U is a T-invariant subspace. Thus, $T|_U$ and T/U are well-defined. It is clear that the matrix form of $T|_U$ is the 1×1 matrix $\begin{pmatrix}\lambda_1\end{pmatrix}$.

We know that $T/U \in \mathcal{L}(V/U)$ and $\dim V/U = \dim V - 1 = n-1 < n$. By the inductive hypothesis, there exists a basis $w_1, w_2, \ldots, w_{n-1}$ of V/U such that $\mathcal{M}(T/U)$ is upper-triangular. Then, recall that V is related to V/U by the quotient map π (see Definition 12.7). Since π is surjective, for each $w_i \in V/U$, there exists some $y_i \in V$ such that $\pi(y_i) = w_i$ for all $i = 1, \ldots, n-1$.

Now, we claim $v_1, y_1, y_2, \ldots, y_{n-1}$ is a basis of V. Since the number of vectors is equivalent to $\dim V$, it remains to show that $v_1, y_1, y_2, \ldots, y_{n-1}$ is linearly independent. Let $c_1, c_2, \ldots, c_n$ be scalars such that

$$c_1v_1 + c_2y_1 + \cdots + c_ny_{n-1} = 0.$$

Since $v_1 \in U$, it follows that $\pi(v_1) = 0$. Then, applying π to both sides gives

$$\pi(c_1v_1 + c_2y_1 + \cdots + c_ny_{n-1}) = c_2w_1 + \cdots + c_nw_{n-1} = 0.$$

Since $w_1, \ldots, w_{n-1}$ is a basis, it follows that $c_2 = \cdots = c_n = 0$. It remains that $c_1v_1 = 0$, so $c_1 = 0$. Thus, $v_1, y_1, y_2, \ldots, y_{n-1}$ is linearly independent.

Now, consider the matrix of T under the basis $v_1, y_1, y_2, \ldots, y_{n-1}$. It is clear that

$$\mathcal{M}(T) = \left(\begin{array}{c|ccc} \lambda_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & \mathcal{M}(T/U, (w_1, \ldots, w_{n-1})) & \\ 0 & & & \end{array}\right),$$

Note that $w_i = y_i + U$ for all $i = 1, \ldots, n-1$, which is why the lower-right block of $\mathcal{M}(T)$ is equivalent to $\mathcal{M}(T/U)$. By the inductive hypothesis $\mathcal{M}(V/U)$ is upper-triangular, so $\mathcal{M}(T)$ is upper-triangular, as desired. □

Last lecture, we proved Proposition 12.11. Now, we will prove some sort of converse.

Corollary 13.6
Suppose A is an $n\times n$ matrix with entries in $\mathbb{C}$. Then, there exists an invertible $n\times n$ matrix S such that SAS^{-1} is upper-triangular.

Note that matrix SAS^{-1} represents the same operator as matrix A, but under a different basis determined by S.[23] [24]

13.3 Existence of Eigenvalues (Real Vector Spaces)

We will continue our discussion of eigenvalues, this time over $\mathbb{R}$.

[23] The operation SAS^{-1} is called *conjugation* by S on matrix A.

[24] In particular, if A is the matrix of T under the basis $v_1, \ldots, v_n$, then SAS^{-1} is the matrix of T under the basis $w_1, \ldots, w_n$ where $v_i = \sum_{j=1}^{n} S_{j,i}w_j$ for $i = 1, 2, \ldots, n$.

Example 13.7
Consider the 2×2 matrix with real entries

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

Suppose $T \in \mathcal{L}(\mathbb{R}^2)$ is the linear operator corresponding to matrix A.

Solving for the eigenvalues of A, we get the equation

$$(\lambda - 0)(\lambda - 0) - 1(-1) = \lambda^2 + 1 = 0.$$

This clearly does not have any real roots, so T has no real eigenvalues.[a]

[a] Note that if $T \in \mathcal{L}(\mathbb{C}^2)$, then T would have eigenvalues i and $-i$.

13.4 Eigenspaces

Now, suppose we have a matrix A with change of basis matrix S such that

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

By Theorem 12.11, $\lambda = 2, 3$ are the eigenvalues of SAS^{-1}. However, both A and SAS^{-1} are matrix representations of the same operator T, so $\lambda = 2, 3$ are also the eigenvalues of A.

This implies that any upper-triangular matrix representation of T must have 2 and 3 on the diagonal.

Guiding Question
Consider the same matrix A as above. It is possible to have a change of basis matrix S_1 such that

$$S_1 A S_1^{-1} = \begin{pmatrix} 2 & * & * \\ 0 & 2 & * \\ 0 & 0 & 3 \end{pmatrix}?$$

In general, is it possible for SAS^{-1} and $S_1 A S_1^{-1}$ to have the same set of eigenvalues along the diagonal, but they appear different number of times?

Answer. *No! We will learn later in the course that number of times each eigenvalue appears on the diagonal is related to the linear operator T, and thus is invariant under conjugation by S and S_1.*

Today, we will discuss a special case of the question above, which is when A is diagonal.

Guiding Question
Does there exist some matrix S such that

$$S \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix} S^{-1} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}?$$

To answer this question, we will first need the following definition.

Definition 13.8 (eigenspace)
Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$. Then, the **eigenspace** of T corresponding to eigenvalue λ is defined by

$$E(\lambda, T) = \{v \in V | T(v) = \lambda v\}.$$

In other words, $E(\lambda, T)$ is the set of all eigenvalues of T corresponding to eigenvalue λ and the 0 vector.

Note that the 0 vector is included in the eigenspace, so $E(\lambda, T)$ is a subspace. This can be shown by seeing that the condition $T(v) = \lambda v$ implies that $(T - \lambda \mathrm{id}_V)(v) = 0$, so

$$E(\lambda, T) = \mathrm{Null}(T - \lambda \mathrm{id}_V).$$

Since the null space of any linear map is a subspace, it follows that $E(\lambda, T)$ is a subspace of V.

Now, we will use eigenspaces to answer the Guiding Question. We will reformulate the Guiding Question in terms of linear operators instead of matrices.

Guiding Question

Suppose $T \in \mathcal{L}(\mathbb{F}^3)$ such that

$$\mathcal{M}(T, \{e_1, e_2, e_3\}) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix},$$

where $\{e_1, e_2, e_3\}$ is the standard basis of $\mathbb{F}^3$. Does there exist a basis v_1, v_2, v_3 of $\mathbb{F}^3$ such that

$$\mathcal{M}(T, \{v_1, v_2, v_3\}) = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}?$$

We will compute $E(2, T)$ with respect to each of the different bases. First, by the matrix form of T with respect to the standard basis, we know that

$$\begin{aligned} Te_1 &= 2e_1 \\ Te_2 &= 3e_2 \\ Te_3 &= 3e_3. \end{aligned}$$

Then, it is clear that

$$\begin{aligned} (T - 2\mathrm{id})e_1 &= 0 \\ (T - 2\mathrm{id})e_2 &= e_2 \\ (T - 2\mathrm{id})e_2 &= e_3. \end{aligned}$$

It follows that $\mathrm{Null}(T - 2\mathrm{id})$ is spanned by e_1, so $E(2, T) = \mathrm{span}(e_1)$. Thus, $E(2, T)$ is a 1-dimensional subspace.

Now, by similar logic as above, the matrix form of T with respect to $\{v_1, v_2, v_3\}$ implies that $E(2, T) = \mathrm{span}(v_1, v_2)$. However, this implies that $E(2, T)$ is a 2-dimensional subspace, which is a contradiction. Therefore, it is impossible to have two different bases where the eigenvalues of T appear with different multiplicities.

In general, we can make the following statement.

Fact 13.9

Suppose

$$\mathcal{M}(T, \{v_1, \dots, v_n\}) = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}.$$

Then, $E(\lambda_T) = \mathrm{span}(\{v_i | \lambda_i = \lambda\})$.

Note that if λ does not appear on the diagonal, then $E(\lambda, T) = \{0\}$, so λ is not an eigenvalue of T. Additionally, it is clear from this statement that $\dim E(\lambda, T)$ is equivalent to the number of times λ appears among $\lambda_1, \lambda_2, \dots, \lambda_n$. Thus, while the order of the diagonal entries $\lambda_1, l_2, \dots, l_n$ may vary depending on the basis, the multiplicities of each value are always the same.

Finally, while we have only shown Fact 13.9 to be true for diagonal matrices, it can be shown this statement is also true for upper-triangular matrices (proof covered in a later lecture).

Recall that if $\lambda_1, \lambda_2, \dots, \lambda_m$ are distinct eigenvalues, then the corresponding eigenvalues $v_1, v_2, \dots, v_m$ are linearly independent.

Corollary 13.10
Let $\{\lambda_1, \ldots, \lambda_m\}$ be the set of eigenvalues of $T \in \mathcal{L}(V)$. Then,

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

is a direct sum.

Proof. To prove that this is a direct sum, we must show that for $v_i \in E(\lambda_i, T)$ such that $v_1 + v_2 + \cdots + v_m = 0$, then all $v_i = 0$. However, since $\lambda_1, \lambda_2, \ldots, \lambda_m$ are distinct, it follows that $v_1, v_2, \ldots, v_m$ are linearly independent, so all $v_i = 0$. □

13.5 Diagonalizable Operators

Consider the following example.

Example 13.11
Suppose $T \in \mathcal{L}(\mathbb{F}^2)$ such that

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

Then, 0 is the only eigenvalue of T. It follows that $E(0, T) = \text{Null}(T) = \text{span}(e_1)$. Thus, 0 being the only eigenvalue of T does not necessarily imply that $E(0, T)$ spans the entire vector space $\mathbb{F}^2$.

In fact, Corollary 13.10 implies that

$$\dim E(\lambda_1, T) + \dim E(\lambda_2, T) + \cdots + \dim E(\lambda_m, T) \leq \dim V,$$

since $E(\lambda_i, T)$ are all subspaces of V.

Now, we are ready to make a new definition.

Definition 13.12
An operator $T \in \mathcal{L}(V)$ is **diagonalizable** if $\mathcal{M}(T)$ under some basis of V is diagonal.

Proposition 13.13
Suppose $T \in \mathcal{L}(V)$ and $\{\lambda_1, \lambda_2, \ldots, \lambda_m\}$ are the eigenvalues of T. Then, the following are equivalent:

1. T is diagonalizable;
2. V has a basis consisting of eigenvectors;
3. $E(\lambda_1, T) \oplus E(\lambda_2, T) \oplus \cdots \oplus E(\lambda_m, T) = V$;
4. $\sum_{i=1}^m E(\lambda_i, T) = \dim V$.

Proof. First, we still show that (3) and (4) are equivalent. In the general case, we know that

$$E(\lambda_1, T) \oplus E(\lambda_2, T) \oplus \cdots \oplus E(\lambda_m, T) \subset V.$$

Then, if $\sum_{i=1}^m E(\lambda_i, T) = \dim V$, it is clear that the inclusion above must become the equality $E(\lambda_1, T) \oplus E(\lambda_2, T) \oplus \cdots \oplus E(\lambda_m, T) = V$. Similarly, in the general case,

$$\sum_{i=1}^m E(\lambda_i, T) \leq \dim V.$$

Then, if $E(\lambda_1, T) \oplus E(\lambda_2, T) \oplus \cdots \oplus E(\lambda_m, T) = V$, it is clear that the inequality above must become the equality $\sum_{i=1}^m E(\lambda_i, T) = \dim V$. Thus, (3) and (4) are equivalent.

Now, we will show that (1) implies (2). Suppose T is diagonalizable. Then, there exists a basis $v_1, v_2, \dots, v_n$ of V such that

$$\mathcal{M}(T) = \begin{pmatrix} a_1 & & & \\ & a_2 & & \\ & & \ddots & \\ & & & a_n \end{pmatrix}.$$

This implies that $Tv_i = a_i v_i$, so each v_i is an eigenvector with eigenvalue a_i. Therefore, $v_1, v_2, \dots, v_n$ is a basis of of V consisting of eigenvectors.

Now, we will show that (2) implies (3). Suppose we have a basis $v_1, v_2, \dots, v_n$ of V consisting of eigenvectors. Then, each $v_i \in E(\lambda_j, T)$ for some j. It follows that all basis vectors

$$v_1, v_2, \dots, v_n \in E(\lambda_1, T) \oplus E(\lambda_2, T) \oplus \cdots \oplus E(\lambda_m, T),$$

so $\operatorname{span}(v_1, v_2, \dots, v_n) = V \subset E(\lambda_1, T) \oplus E(\lambda_2, T) \oplus \cdots \oplus E(\lambda_m, T)$. Furthermore, by the definition of sum of subspaces, we know that $E(\lambda_1, T) \oplus E(\lambda_2, T) \oplus \cdots \oplus E(\lambda_m, T) \subset V$. Therefore, $E(\lambda_1, T) \oplus E(\lambda_2, T) \oplus \cdots \oplus E(\lambda_m, T) = V$.

Lastly, we will show that (3) implies (1). Suppose $E(\lambda_1, T) \oplus E(\lambda_2, T) \oplus \cdots \oplus E(\lambda_m, T) = V$. Let $\dim E(\lambda_i, T) = d_i$. Then, choose a basis $v_{i,1}, v_{i,2}, \dots, v_{i,d_i}$ of $E(\lambda_i, T)$ for $i = 1, 2, \dots, m$. It follows that

$$v_{1,1}, v_{1,2}, \dots, v_{1,d_1}, v_{2,1}, v_{2,2}, \dots, v_{2,d_2}, \dots, \dots, v_{m,1}, v_{m_2}, \dots, v_{m,d_m}$$

is a basis of V. Under this basis, $\mathcal{M}(T)$ becomes the diagonal matrix

$$\mathcal{M}(T) = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{pmatrix},$$

where the big λ_i represents the diagonal matrix of size $d_i \times d_i$ with all diagonal entries λ_i (in other words, $\mathcal{M}(T)$ is the diagonal matrix where each λ_i appears exactly d_i times). Thus, T is diagonalizable. $\square$

14 Inner Product Spaces

14.1 Review

Last time, we discussed the notion of eigenspaces, where for any $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{F}$, the eigenspace of T corresponding to λ is defined as $E(\lambda, T) = \{v \in V | Tv = \lambda v\}$. Furthermore, if $\{\lambda_1, \dots, \lambda_m\}$ are all eigenvalues of T, then $E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T) \subset V$. This also implies that $\sum_{i=1}^m \dim E(\lambda_i, T) \leq \dim V$.

14.2 Diagonal Operators (continued)

In this section, we will wrap-up our discussion of diagonal operators from last lecture. Note that if λ is an eigenvalue of T, then $\dim E(\lambda, T) \geq 1$. So $\sum_{i=1}^m \dim E(\lambda_i, T) \geq m$.

Theorem 14.1
Suppose $T \in \mathcal{L}(V)$ and T has $\dim V$ distinct eigenvalues. Then, T is diagonalizable.

Proof. Let $n = \dim V$. Then,

$$\sum_{i=1}^m \dim E(\lambda_i, T) \geq \dim V,$$

since $\dim E(\lambda_i, T) \geq 1$ for all λ_i. By the definition of direct sum, we know that

$$\sum_{i=1}^m \dim E(\lambda_i, T) \leq \dim V.$$

It follows that $\sum_{i=1}^m \dim E(\lambda_i, T) = \dim V$, which implies that T is diagonaliazble by Theoream 13.13. □

Example 14.2
Suppose $T \in \mathcal{L}(V)$ has corresponding matrix

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

Then, T has 3 distinct eigenvalues $\lambda = 1, 2, 3$, so T is diagonalizable.

14.3 Inner Product Spaces

Previously, we have been considering abstract vector spaces, which are sets of vectors with abstract addition and scalar multiplication. Now, inner product spaces are vector spaces with one piece of extra structure: namely, the notion of the *length* (also called the *norm*) of a vector.

Before we give a definition of an inner product, let's look at two standard examples.

Example 14.3
Suppose $V = \mathbb{R}^n$. Then, the ***dot product*** of vectors $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ is defined by

$$x \cdot y = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

Note that $x \cdot y \in \mathbb{R}$. The dot product is also known as the Euclidean product.

The norm of x (denoted as $||x||$) is defined by

$$||x|| = \sqrt{x \cdot x} = \sqrt{x_1^2 + x_2^2 + \cdots x_n^2}.$$

From our knowledge of $\mathbb{R}^n$, it is clear that this definition of the norm is measuring the length of x.

Example 14.4
Suppose $V = \mathbb{C}^n$. Then, $\langle x, y\rangle$ for vectors $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ is defined by

$$\langle x, y\rangle = x_1\overline{y_1} + x_2\overline{y_2} + \cdots + x_n\overline{y_n}.$$

Note that $\langle x, y\rangle \in \mathbb{C}$. This product is also known as the Hermitian product.

The norm of x is similarly defined by

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

Recall that for a complex number $z = a + bi$, then $z\overline{z} = |z|^2 = a^2 + b^2$. So,

$$\langle x, x\rangle = x_1\overline{x_1} + \cdots + x_n\overline{x_n} = |x_1|^2 + \cdots |x_n|^2 \geq 0$$

Thus, $\langle x, x\rangle$ is a nonnegative real number, so $||x||$ is well-defined.

Now, we will give more evidence to show that this definition of the norm is measuring the length of x. Let $c \in \mathbb{C}$. Then,

$$\begin{aligned}\langle cx, cx\rangle &= |cx_1|^2 + \cdots + |cx_n|^2 \\ &= |c|^2(|x_1|^2 + \cdots + |x_n|^2) \\ &= |c|^2\langle x, x\rangle.\end{aligned}$$

Thus, $||cx|| = \sqrt{\langle cx, cx\rangle} = |c|\sqrt{\langle x, x\rangle} = |c|\,||x||$. Therefore, if some vector is scaled by a factor of c, then its norm (or length) is scaled by a factor of $|c|$.

Now, we are ready to give the formal definition of the inner product. First, we will consider inner products over real vector spaces.

Definition 14.5 (inner product over $\mathbb{R}$-vector space)
Suppose V is a $\mathbb{R}$-vector space. An **inner product** on V is a function that takes each ordered pair (x, y) of vectors in V to the number $\langle x, y\rangle \in \mathbb{R}$ and satisfies the following conditions:

- additivity in the first variable: $\langle x + x', y\rangle = \langle x, y\rangle + \langle x', y\rangle$
- homogeneity in the first variable: $\langle cx, y\rangle = c\langle x, y\rangle$ for all $c \in \mathbb{R}$
- symmetry: $\langle x, y\rangle = \langle y, x\rangle$
- positivity: $\langle x, x\rangle \geq 0$
- definiteness: $\langle x, x\rangle = 0$ if and only if $x = 0$

Let's verify that the dot product on $\mathbb{R}^n$ is an inner product.

Example 14.6
Suppose $\langle x, y\rangle = x \cdot y = x_1y_1 + \cdots + x_ny + n$. We must verify all the necessary conditions:

- additivity in the first variable: $(x + x') \cdot y = (x_1 + x_1')y_1 + \cdots + (x_n + x_n')y_n = (x_1y_1 + \cdots + x_ny_n) + (x_1'y_1 + \cdots + x_n'y_n) = x \cdot y + x' \cdot y$
- homogeneity in the first variable: $(cx) \cdot y = cx_1y_1 + \cdots + cx_ny_n = c(x \cdot y)$
- symmetry: $x \cdot y = x_1y_1 + \cdots + x_ny_n = y_1x_1 + \cdots + y_nx_ny \cdot x$
- positivity: $x \cdot x = x_1^2 + \cdots + x_n^2 \geq 0$
- definiteness: $x \cdot x = x_1^2 + \cdots + x_n^2 = 0 \iff x_1 = \cdots = x_n = 0 \iff x = 0$

Thus, $x \cdot y$ is an inner product on $\mathbb{R}^n$.

We can use the notation $V, \langle\cdot,\cdot\rangle$ to represent a vector space V with inner product $\langle\cdot,\cdot\rangle$.

By the definition of an inner product, $\langle \cdot, \cdot \rangle$ is additive and homogeneous in the first variable.

Guiding Question
Is $\langle \cdot, \cdot \rangle$ also additive and homogeneous in the second variable?

Answer. *Yes! We know that*

$$\begin{aligned} \langle x, y+z \rangle &= \langle y+z, x \rangle \\ &= \langle y, x \rangle + \langle z, x \rangle \\ &= \langle x, y \rangle + \langle x, z \rangle, \end{aligned}$$

where the second equality follows from additivity in the first variable and the first and third inequalities follow from symmetry. Thus, $\langle \cdot, \cdot \rangle$ is additive in the second variable. The proof to show that $\langle \cdot, \cdot \rangle$ is homogeneous in the second variable is identical.[25]

Now, we will define inner products over complex vector spaces.

Definition 14.7 (inner product over $\mathbb{C}$-vector space)
Suppose V is a $\mathbb{C}$-vector space. An **inner product** on V is a function that takes each ordered pair (x, y) of vectors in V to the number $\langle x, y \rangle \in \mathbb{C}$ and satisfies the following conditions:

- additivity in the first variable: $\langle x + x', y \rangle = \langle x, y \rangle + \langle x', y \rangle$
- homogeneity in the first variable: $\langle cx, y \rangle = c\langle x, y \rangle$ for all $c \in \mathbb{C}$
- conjugate symmetry: $\langle x, y \rangle = \overline{\langle y, x \rangle}$
- positivity: $\langle x, x \rangle \geq 0$
- definiteness: $\langle x, x \rangle = 0$ if and only if $x = 0$

Note that the main difference between inner products in real and complex vector spaces is the third condition (*conjugate symmetry* instead of *symmetry*).

Example 14.8
Let's verify that $\langle x, y \rangle = x_1\overline{y_1} + \cdots + x_n\overline{y_n}$ is an inner product in $\mathbb{C}^n$. We know that

$$\begin{aligned} \langle x, y \rangle &= x_1\overline{y_1} + \cdots + x_n\overline{y_n} \\ &= \overline{\overline{x_1}y_1 + \cdots + \overline{x_n}y_n} \\ &= \overline{\langle y, x \rangle}, \end{aligned}$$

so $\langle \cdot, \cdot \rangle$ satisfies conjugate symmetry. The verification of the rest of the conditions is analogous to Example 14.6.

Recall that real inner products are also additive and homogeneous over the second variable.

Guiding Question
Are complex vector spaces also additive and homogeneous over the second variable?

Answer. *Sort of! Complex inner products are additive over the second variable, the proof of which is identical to the real case. However,*

$$\begin{aligned} \langle x, cy \rangle &= \overline{\langle cy, x \rangle} \\ &= \overline{c}\overline{\langle y, x \rangle} \\ &= \overline{c}\langle x, y \rangle, \end{aligned}$$

[25] We say that $\langle \cdot, \cdot \rangle$ is *bilinear*, meaning that $\langle \cdot, \cdot \rangle$ is linear in both variables.

where the second equality follows from homogeneity over the second variable and the first and third equalities follows from conjugate symmetry. Thus, $\langle \cdot, \cdot \rangle$ *has* conjugate homogeneity *over the second variable.*[26]

So far, we have only looked at our two standard examples of inner product spaces. Now, we will give more examples.

Example 14.9
First, we will look at a slight variation of our standard example. Suppose $V = \mathbb{C}^n$ and define

$$\langle x, y \rangle = c_1 x_1 \overline{y_1} + \cdots + c_n x_n \overline{y_n}$$

for $c_1, \ldots, c_n \in \mathbb{C}$. Under what values of $c_1, \ldots, c_n$ is $\langle \cdot, \cdot \rangle$ an inner product space?

To answer this question, we must check the five conditions in Definition 14.7.

- Additivity and homogeneity: It is easy to verify that $\langle \cdot, \cdot \rangle$ satisfies additivity and homogeneity in the first slot for any $c_1, \ldots, c_n$.
- Conjugate symmetry: Note that

$$\begin{aligned} \langle x, y \rangle &= \overline{\langle y, x \rangle} \\ &= \overline{c_1 y_1 \overline{x_1} + \cdots + c_n y_n \overline{x_n}} \\ &= \overline{c_1} x_1 \overline{y_1} + \cdots + \overline{c_n} x_n \overline{y_n}. \end{aligned}$$

  So, to satisfy conjugate symmetry, $c_i = \overline{c_i}$ for all c_i. In other words, all c_i must be real.
- Positivity: Note that

$$\langle x, x \rangle = c_1 |x_1|^2 + \cdots + c_n |x_n|^2 \geq 0.$$

  Since x is arbitrary, we can choose x to be the vector such that $x_i = 1$ and all other entries are 0, which implies that all $c_i \geq 0$.
- Definiteness: If any $c_i = 0$, then let x be the vector such that $x_i = 1$ and all other entries are 0. Then

$$\langle x, x \rangle = c_1 \cdot 0 + \cdots + c_i \cdot 1 + \cdots + c_n \cdot 0 = c_i = 0,$$

  which violates definiteness. Thus, all $c_i > 0$.

In conclusion, $\langle x, y \rangle = c_1 x_1 \overline{y_1} + \cdots + c_n x_n \overline{y_n}$ is an inner product if $c_1, \ldots, c_n$ are positive real numbers.

Example 14.10
Suppose $V = \mathbb{R}^n$ and define

$$\langle x, y \rangle = c_1 x_1 y_1 + \cdots + c_n x_n y_n$$

for $c_1, \ldots, c_n \in \mathbb{R}$. By reasoning identical to the logic in Example 14.9, $\langle \cdot, \cdot \rangle$ is an inner product when $c_1, \ldots, c_n$ are positive.

[26] We say that $\langle \cdot, \cdot \rangle$ is *sesquilinear* (the prefix *sesqui* meaning "one and a half"), meaning that $\langle \cdot, \cdot \rangle$ is linear in the first variable, but only additive and conjugate homogeneous in the second variable.

Example 14.11
Suppose $V = \mathbb{C}^2$ and define
$$\langle x, y\rangle = x_1\overline{y_2} + x_2\overline{y_1}.$$
Then,
$$\langle x, x\rangle = x_1\overline{x_2} + x_2\overline{x_1} = 2\,\mathrm{Re}(x_1\overline{x_2}),$$
where the last equality follows from that $x_1\overline{x_2} = \overline{x_2\overline{x_1}}$. Since x_1, x_2 are arbitrary, $x_1\overline{x_2}$ can essentially be any complex number, in particular a number with a negative real part, which violates positivity. For instance, if $x = (1, -1)$, then
$$\langle x, x\rangle = (1)(-1) + (-1)(1) = -2 < 0.$$
Thus, $\langle\cdot,\cdot\rangle$ is not a valid inner product space.

Furthermore, if $x_1\overline{x_2}$ is purely imaginary, then $\langle x, x\rangle = 2\,\mathrm{Re}(x_1\overline{x_2}) = 0$ where x is not necessarily 0. Thus, $\langle\cdot,\cdot\rangle$ also violates definiteness.

Before we move onto the next section, we will need the following definition. Recall that in Example 14.4, we defined the norm of x as $||x|| = \sqrt{\langle x, x\rangle}$. Now, we will expand this definition to the general case.

Definition 14.12 (norm)
Suppose $V, \langle\cdot,\cdot\rangle$ is any inner product space. Then, for $x \in V$, the **norm** of x is given by
$$||x|| = \sqrt{\langle x, x\rangle}.$$

14.4 Orthogonality

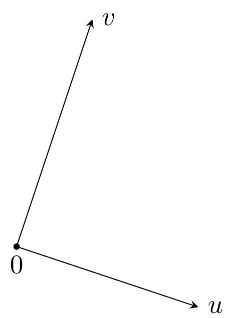
Let $V, \langle\cdot,\cdot\rangle$ be an inner product space.

Definition 14.13 (orthogonal)
Let vectors $u, v \in V$. Then, u is **orthogonal** to v if $\langle u, v\rangle = 0$.[a]

[a] The notation $u \perp v$ is used to mean "u is orthogonal to v."

To better understand this definition, we will first show that this definition of orthogonal vectors makes sense in $\mathbb{R}^2$.

Suppose $V = \mathbb{R}^2$ and $\langle\cdot,\cdot\rangle$ is the dot product. Let $u, v \in V$ be nonzero vectors as seen in the diagram below.



We will show that $u \cdot v = 0$ if and only if the angle between u and v is 90°. To prove this, we will first need the following lemma.

Lemma 14.14
Suppose $V, \langle\cdot,\cdot\rangle$ is any inner product space. If $\langle u, v\rangle = 0$, then
$$||u + v||^2 = ||u||^2 + ||v||^2.$$[a]

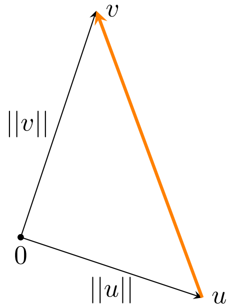
[a] In Axler, this is referred to as the Pythagorean Theorem.

Proof. It follows that

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

where the last equality follows from $\langle u,v\rangle = \langle v,u\rangle = 0$. □

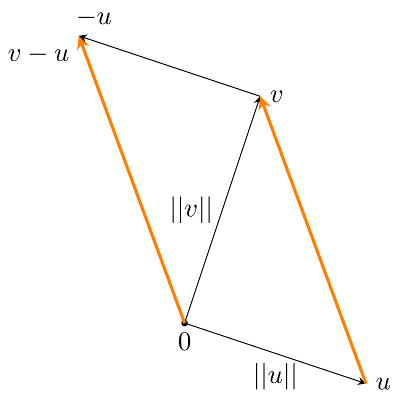
Now, consider the triangle in the diagram below.



It is clear that the two legs of the triangle have lengths $||u||$ and $||v||$.

Guiding Question
How can we describe the orange vector in terms of vectors u and v?

Answer. *One way to find the answer is to see this is to translate the orange vector so that it begins at the origin.*



From this, it is clear that the orange vector is $v-u$.

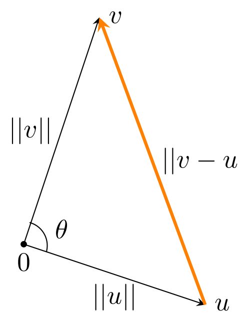
Thus, the triangle has side lengths $||u||, ||v||$, and $||v-u||$. Suppose that $u \cdot v = 0$. Then, it follows that $(-u)\cdot v = -(u\cdot v) = 0$. By Lemma 14.14,

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

Finally, we know that $||-u|| = ||u||$, so

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

Now, let the angle between u and v have degree measure θ.



By the Law of Cosines,

$$\cos\theta = \frac{||v||^2 + ||u||^2 - ||v-u||^2}{2||v||\,||u||}.$$

However, we know that $||v||^2 + ||u||^2 - ||v-u||^2 = 0$, so $\cos\theta = 0$. Therefore, $\theta = 90°$, as desired.

Thus, our definition of orthogonality in Definition 14.13 makes sense with the usual sense of orthogonality in $\mathbb{R}^2$.

14.5 Cauchy-Schwarz Inequality

Now, we will introduce the Cauchy-Schwarz Inequality.

Theorem 14.15 (Cauchy-Schwarz Inequality)
Suppose $V, \langle\cdot,\cdot\rangle$ is an inner product space and $u, v \in V$. Then,

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

We will leave the proof of this for next lecture. First, we will use the two standard examples from earlier to give us a better understand of what the Cauchy-Schwarz Inequality is telling us.

Example 14.16
Suppose $V = \mathbb{R}^n$ and $\langle\cdot,\cdot\rangle$ is the dot product. Let $x, y \in V$. By the Cauchy-Schwarz Inequality,

$$|x \cdot y| \leq ||x||\,||y||.$$

Squaring both sides and expanding out the dot product gives

$$(x_1y_1 + \cdots + x_ny_n)^2 \leq (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2).$$

A special case of this result is given by taking $y_1 = \cdots = y_n = 1$, which gives

$$(x_1 + \cdots + x_n)^2 \leq n(x_1^2 + \cdots + x_n^2).$$[a]

A familiar example of the above inequality is the $n = 2$ case,

$$(x_1 + x_2)^2 \leq 2(x_1^2 + x_2^2).$$

Expanding out both sides and moving all the terms to one side gives

$$0 \leq (x_1 - x_2)^2,$$

which is the trivial inequality.[b]

[a] This kind of inequality is frequently used in calculus or mathematical analysis in order to approximate certain quantities.
[b] It is also possible to prove the general case $(x_1 + \cdots + x_n)^2 \leq n(x_1^2 + \cdots + x_n^2)$ in a similar fashion to the $n = 2$ case using the trivial inequality.

Example 14.17
Suppose $V = \mathbb{C}^n$ and $\langle \cdot, \cdot \rangle$ is defined as in Example 14.4. Let $x, y \in V$. Then, squaring both sides of the Cauchy-Schwarz Inequality gives us

$$|x_1\overline{y_1} + \cdots + x_n\overline{y_n}|^2 \leq (|x_1|^2 + \cdots + |x_n|^2)(|y_1|^2 + \cdots + |y_n|^2).$$

15 Cauchy-Schwarz Inequality and Orthonormal Bases

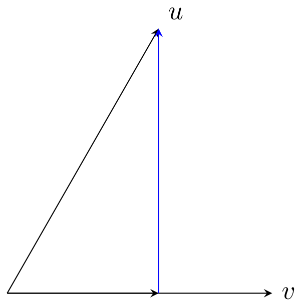
15.1 Review

Last time, we introduced the inner product space and ended with the Cauchy-Schwarz Inequality. Recall Theorem 14.15 for the statement of the Cauchy-Schwarz Inequality.

Before we prove the Cauchy-Schwarz inequality, we first need the notion of orthogonal decomposition.

15.2 Orthogonal Decomposition

Suppose we have two vectors u and v in the two-dimensional real plane. We wish to decompose u into the sum of two vectors, one which is parallel to v and one which is perpendicular to v.



In the general case, we want to express u as

$$u = cv + w$$

where $c \in \mathbb{F}$ and w is orthogonal to v.

If $v = 0$, then the orthogonal decomposition of u is simply w, since all vectors are orthogonal to the zero vector.

Now, assume $v \neq 0$. Suppose we have the decomposition $u = cv + w$ with w orthogonal to v. Then,

$$\begin{aligned}\langle u, v\rangle &= \langle cv + w, v\rangle \\ &= c\langle v, v\rangle + \langle w, v\rangle \\ &= c\langle v, v\rangle.\end{aligned}$$

It follows that $c = \frac{\langle u,v\rangle}{\langle v,v\rangle}$ and $w = u - \frac{\langle u,v\rangle}{\langle v,v\rangle}v$. We can also verify that our expression for w is indeed orthogonal to v :

$$\begin{aligned}\langle w, v\rangle &= \langle u - \tfrac{\langle u,v\rangle}{\langle v,v\rangle}v, v\rangle \\ &= \langle u, v\rangle - \tfrac{\langle u,v\rangle}{\langle v,v\rangle}\langle v, v\rangle \\ &= 0.\end{aligned}$$

Theorem 15.1 (orthogonal decomposition)
Suppose $u, v \in V$ with $v \neq 0$. Let $c = \frac{\langle u,v\rangle}{\langle v,v\rangle}$ and $w = u - \frac{\langle u,v\rangle}{\langle v,v\rangle}v$. Then, u can be expressed as

$$u = cv + w$$

such that w is orthogonal to v.

Now, we are ready to prove the Cauchy-Schwarz Inequality.

15.3 Cauchy-Schwarz Inequality (continued)

As a sidenote, note that the statement of the Cauchy-Schwarz Inequality does not require V to be finite-dimensional. This is intuitive because the Cauchy-Schwarz Inequality only involves two vectors u and v, so you can always restrict your attention to the subspace spanned by u and v. Later, we will look at examples of infinite-dimensional inner product spaces, and Cauchy-Schwarz will still have applications there.

Now, we will prove the Cauchy-Schwarz Inequality.

Proof of Cauchy-Schwarz Inequality (Theorem 14.15). If $v = 0$, then both sides of the desired inequality equal 0, and we are done. Thus, assume $v \neq 0$. Consider the orthogonal decomposition

$$u = cv + w$$

as described in Theorem 15.1. Then,

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

since $\langle w, v\rangle = 0$. Furthermore,

$$\begin{aligned} ||u|| &= \sqrt{\langle u, u\rangle} \\ &= \sqrt{\langle cv + w, cv + w\rangle} \\ &= \sqrt{\langle cv, cv\rangle + \langle cv, w\rangle + \langle w, cv\rangle + \langle w, w\rangle} \\ &= \sqrt{\langle cv, cv\rangle + \langle w, w\rangle} \\ &= \sqrt{||cv||^2 + ||w||^2} \\ &\geq ||cv|| \\ &= |c|\,||v||. \end{aligned}$$

It follows that

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

which gives the desired inequality. ☐

Guiding Question
When does equality hold in the Cauchy-Schwarz Inequality?

Answer. *Equality holds when* $\sqrt{||cv||^2 + ||w||^2} = ||cv||$, *so* $w = 0$. *This implies that* $u = cv$, *so* u *is a scalar multiple of* v.

Conversely, if $u = cv$ *for some* $c \in \mathbb{C}$, *then*

$$\langle u, v\rangle = |c|\langle v, v\rangle = |c|\,||v||^2 = ||u||\,||v||.$$

Therefore, equality holds if and only if u *is a scalar multiple of* v. *Note that this includes the case where* u *or* v *is the zero vector.*

Now, we will look at an infinite-dimensional example of the Cauchy-Schwarz Inequality.

Example 15.2
Suppose V be the set of all continuous functions $f : [-1, 1] \to \mathbb{R}$. Note that V contains all real-valued polynomials, so V must be infinite-dimensional (since $\mathcal{P}(\mathbb{R})$ is already infinite-dimensional).

Let $f, g \in V$. Then

$$\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)\,dx$$

is an inner product on V. It is clear that additivity, homogeneity, and symmetry hold. We know that

$$||f|| = \sqrt{\langle f, f\rangle} = \left(\int_{-1}^{1} (f(x))^2\,dx\right)^{\frac{1}{2}}.$$

Thus, positivity holds because $f(x)^2 \geq 0$. Finally, it follows that $||f|| = 0$ if and only if $f = 0$, so definiteness holds.

Now, applying Cauchy-Schwarz to f, g and squaring both sides gives

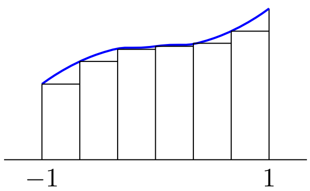
$$\left(\int_{-1}^{1} f(x)g(x)\,dx\right)^2 \leq \left(\int_{-1}^{1} (f(x))^2\right)\left(\int_{-1}^{1} (g(x))^2\right).$$

Guiding Question
Compare the above inequality with the following inequality from Example 14.16:

$$(x_1y_1 + \cdots + x_ny_n)^2 \leq (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2).$$

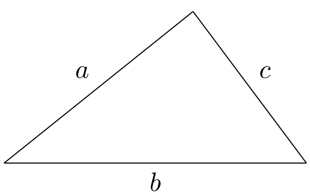
What is similar between these two inequalities?

Answer. *Consider the area under the blue curve in the following diagram:*

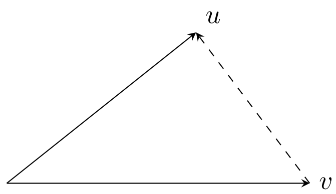


The expression $\int_{-1}^{1} f(x)g(x)\,dx$ can be seen as taking the area using an integral, while the expression $x_1y_1 + \cdots + x_ny_n$ can be seen as taking the area using a Riemann sum. Thus, the inequality from Example 14.15 can be seen as a discrete version of the continuous inequality from Example 15.2. Note that the discrete version of the inequality comes from a finite-dimensional inner product space, while the continuous version comes from an infinite-dimensional inner product space.

Now, we will discuss another important inequality, known as the Triangle Inequality. This inequality comes from the elementary observation that the sum of any two sides of a triangle must be greater than or equal to the third side. In other words, $a + b \geq c$.



If we interpret the above diagram in terms of vectors, we find that $a = ||u||, b = ||v||$, and $c = ||u - v||$.



Thus,

$$||u|| + ||v|| \geq ||u - v||.$$

Since $||v|| = ||-v||$, we can replace v with $-v$ in the above inequality to get the Triangle Inequality:

$$||u|| + ||v|| \geq ||u + v||.$$

While this example only concerns the two-dimensional plane, the Triangle Inequality can be expanded to any inner product space.

Theorem 15.3 (Triangle Inequality)
Suppose $u, v \in V$. Then,

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

Proof. It is equivalent to show that

$$||u + v||^2 \leq (||u|| + ||v||)^2.$$

Then,

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

where the last inequality follows from the Cauchy-Schwarz Inequality. □

15.4 Orthonormal Bases

We will shift our discussion to the topic of orthonormal bases. First, we will need the following definition.

Definition 15.4 (orthonormal)
A list of vectors $v_1, v_2, \ldots, v_n$ is **orthonormal** if

$$\begin{cases} \langle v_i, v_j\rangle = 0 & \text{for } i \neq j, \\ ||v_i|| = 1 & \forall\, i = 1, \ldots, n. \end{cases}$$

Equivalently,

$$\langle v_i, v_j\rangle = \delta_{ij} = \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

Example 15.5
In $\mathbb{R}^n$, the standard basis

$$\begin{aligned} e_1 &= (1, 0, \ldots, 0) \\ &\vdots \\ e_n &= (0, \ldots, 0, 1) \end{aligned}$$

is an orthonormal list.

The folllowing is an important lemma.

Lemma 15.6
Any orthonormal list is linearly independent.

Proof. Suppose $v_1, \ldots, v_n$ be an orthonormal list. Let $c_1, \ldots, c_n$ be scalars such that

$$c_1 v_1 + c_2 v_2 + \cdots + c_n c_n = 0.$$

Taking the inner product of the left-hand side with v_i, we get

$$\begin{aligned} \langle c_1 v_1 + \cdots + c_n v_n, v_i \rangle &= \langle c_1 v_1, v_i \rangle + \cdots + \langle c_i v_i, v_i \rangle + \cdots + \langle c_n, v_n, v_i \rangle \\ &= \langle c_i v_i, v_i \rangle \\ &= c_i. \end{aligned}$$

This has to be equivalent to $\langle 0, v_i \rangle = 0$, so $c_i = 0$ for all $i = 1, 2, \ldots, n$. Therefore, $v_1, \ldots, v_n$ is linearly independent. □

Thus, all orthonormal lists are linearly independent. Therefore, if we have the right number of vectors in an orthonormal list, the list must form a basis.

Definition 15.7 (orthonormal basis)
An **orthonormal** basis of V is an orthonormal list that is also a basis of V.

Equivalently, an orthonormal basis is an orthonormal list of $\dim V$ vectors.

Example 15.8
The standard basis of $\mathbb{R}^n$ is an orthonormal basis.

Example 15.9
Suppose $V = \mathcal{P}_n(\mathbb{R})$ with inner product

$$\langle f, g \rangle = \int_{-1}^{1} f(x) g(x) \, dx.$$

Consider the basis $1, x, x^2, \ldots, x^n$. Then,

$$\langle f_i, f_j \rangle = \int_{-1}^{1} x^i x^j \, dx = \int_{-1}^{1} x^{i+j} \, dx = \begin{cases} 0 & \text{if } i + j \text{ is odd,} \\ \frac{2}{i+j+1} & \text{if } i + j \text{ is even.} \end{cases}$$

Thus, it is clear that $1, x, x^2, \ldots, x^n$ is not an orthonormal basis. For instance,

$$\langle f_1, f_1 \rangle = \frac{2}{3} \neq 1 \quad \text{and} \quad \langle f_0, f_2 \rangle = \frac{2}{3} \neq 0.$$

Now, we will discuss why orthonormal bases are useful.

Theorem 15.10
Suppose $e_1, e_2, \ldots, e_n$ is an orthonormal basis of V and $v \in V$. Then,

$$v = \langle v, e_1\rangle e_1 + \langle v, e_2\rangle e_2 + \cdots + \langle v, e_n\rangle e_n.$$

Proof. Since $e_1, \ldots, e_n$ is a basis, there exists scalars $c_1, \ldots, c_n$ such that

$$v = c_1 e_1 + \cdots + c_n e_n.$$

Then,

$$\begin{aligned}\langle v, e_i\rangle &= \langle c_1 e_1 + \cdots + c_n e_n, e_i\rangle \\ &= c_1\langle e_1, e_i\rangle + \cdots + c_i\langle e_i, e_i\rangle + \cdots + c_n\langle e_n, e_i\rangle \\ &= c_i,\end{aligned}$$

as desired. □

Note that if $e_1, \ldots, e_n$ is not an orthonormal basis, finding coefficients $c_1, \ldots, c_n$ would require solving a system of n equations in n variables. Thus, finding coefficients $c_1, \ldots, c_n$ is much easier with an orthonormal basis.

15.5 Gram-Schmidt Procedure

Now, we will show that every vector space has an orthonormal basis, as well as how to find such an orthonormal basis.

Algorithm 15.11 (Gram-Schmidt Procedure)
Suppose $v_1, \ldots, v_n$ is a basis of V. Then, there exists an orthonormal basis $e_1, \ldots, e_n$ of V such that

$$e_j \in \operatorname{span}(v_1, \ldots, v_j)$$

for $j = 1, \ldots, n$.

We will break down the algorithm into steps:

- **Step 1:** Since $e_1 \in \operatorname{span}(v_1)$, it follows that $e_1 = c_1 v_1$ for some scalar c_1. We want

$$||e_1|| = ||c_1 v_1|| = |c_1|\,||v_1|| = 1,$$

  so $|c_1| = \frac{1}{||v_1||}$. Thus, we can let $c_1 = \frac{1}{v_1}$, so $e_1 = v_1/||v_1||$.[27]

- **Step 2:** We know that $e_2 \in \operatorname{span}(v_1, v_2) = \operatorname{span}(e_1, v_2)$. Let $e_2 = c_1 e_1 + c_2 v_2$. We want

$$\begin{aligned}\langle e_2, e_1\rangle &= \langle c_1 e_1 + c_2 v_2, e_1\rangle \\ &= c_1\langle e_1, e_1\rangle + c_2\langle e_1, v_2\rangle \\ &= c_1 + c_2\langle v_2, e_1\rangle \\ &= 0,\end{aligned}$$

  so $c_1 = -c_2\langle v_2, e_1\rangle$. Then, by taking $c_1 = 1$, it follows that

$$e_2' = v_2 - \langle v_2, e_1\rangle e_1$$

  satisfies $\langle e_2', e_1\rangle = 0$. Now, we also want $\langle e_2, e_2\rangle = 1$, which we can obtain by normalizing e_2'. Thus,

$$e_2 = \frac{e_2'}{||e_2'||} = \frac{v_2 - \langle v_2, e_1\rangle e_1}{||v_2 - \langle v_2, e_1\rangle e_1||}.$$

[27] This procedure is called *normalizing* v_1 (i.e. scaling v_1 to have norm equal to 1).

- **Step 3:** Similarly to Step 2, $\langle e_3', e_1\rangle = \langle e_3', e_2\rangle = 0$. By the same logic as Step 2, it follows that

$$c_1 = -\langle v_3, e_1\rangle \quad \text{and} \quad c_2 = -\langle v_3, e_2\rangle,$$

so $e_3' = v_3 - \langle v_3, e_1\rangle e_1 - \langle v_3, e_2\rangle e_2$. Thus,

$$e_3 = \frac{e_3'}{||e_3'||} = \frac{v_3 - \langle v_3, e_1\rangle e_1 - \langle v_3, e_2\rangle e_2}{||v_3 - \langle v_3, e_1\rangle e_1 - \langle v_3, e_2\rangle e_2||}.$$

Finally, if we repeat this procedure for a total of n steps, the result will be an orthonormal basis $e_1, \dots, e_n$.

Example 15.12
Consider the inner product space $\mathcal{P}_2(\mathbb{R})$ with

$$\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)\,dx.$$

Consider the basis

$$f_0 = 1, f_1 = x, f_2 = x^2.$$

We will use the Gram-Schmidt Procedure to find an orthonormal basis of $\mathcal{P}_2(\mathbb{R})$.

First, recall from Example 15.9 that

$$\langle f_i, f_j\rangle = \begin{cases} 0 & \text{if } i+j \text{ is odd,} \\ \frac{2}{i+j+1} & \text{if } i+j \text{ is even.} \end{cases}$$

- **Step 1:** We know that $\langle f_0, f_0\rangle = \frac{2}{0+0+1} = 2$. Thus,

$$e_0 = \frac{f_0}{||f_0||} = \frac{1}{\sqrt{2}}.$$

- **Step 2:** Recall that $e_1' = f_1 - \langle f_1, e_0\rangle e_0$. We know that $\langle f_1, e_0\rangle = \frac{1}{\sqrt{2}}\langle f_1, f_0\rangle = 0$, so $e_1' = f_1$. Thus,

$$e_1 = \frac{f_1}{||f_1||} = \frac{x}{\frac{2}{1+1+1}} = \sqrt{\frac{3}{2}}x.$$

- **Step 3:** Recall that $e_2' = f_2 - \langle f_2, e_0\rangle e_0 - \langle f_2, e_1\rangle e_1$. We know that

$$\langle f_2, e_0\rangle = \frac{1}{\sqrt{2}}\langle f_2, f_0\rangle = \frac{\sqrt{2}}{3} \quad \text{and} \quad \langle f_2, e_1\rangle = \sqrt{\frac{3}{2}}x\langle f_2, f_1\rangle = 0,$$

so $e_2' = x^2 - \frac{sqrt2}{3} \cdot \frac{1}{\sqrt{2}} = x^2 - \frac{1}{3}$. Then,

$$||e_2'||^2 = \langle e_2', e_2'\rangle = \int_{-1}^{1} \left(x^4 - \frac{2}{3}x + \frac{1}{9}\right)\,dx = \frac{8}{45},$$

so

$$e_2 = \frac{e_2'}{||e_2'||} = \sqrt{\frac{45}{8}}\left(x^2 - \frac{1}{3}\right).$$

Therefore,

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

is an orthonormal basis of $\mathcal{P}_2(\mathbb{R})$.

16 Gram-Schmidt Procedure and Orthogonal Projection

16.1 Review

Last time, we talked about the notion of orthonormal bases (see Definition 15.7) and Gram-Schmidt Procedure (see Algorithm 15.11).

16.2 Matrix Representation of Inner Products

Suppose V is an inner product space over $\mathbb{C}$. Then, the inner product $\langle \cdot, \cdot \rangle$ can be represented as a matrix.

To do this, choose an arbitrary basis $v_1, \ldots, v_n$ of V. Then, let A be the $n \times n$ matrix such that

$$A = \big(\langle v_i, v_j \rangle\big)_{i,j}.$$

If we know $\langle v_i, v_j \rangle$ for all i, j, then we can compute $\langle u, v \rangle$ for all $u, v \in V$. Let

$$u = \sum_{i=1}^{n} a_i v_i \quad \text{and} \quad v = \sum_{i=1}^{n} b_i v_i.$$

Then, by linearity and conjugate linearity of inner products, it follows that

$$\langle u, v \rangle = \sum_{i,j=1}^{n} a_i \overline{b_j} \langle v_i, v_j \rangle.$$

Thus, the inner product of any two vectors in V are completely determined by the entries of A.

However, the matrix A are not arbitrary: in particular, A must satisfy some specific properties:

- By conjugate symmetry, $\langle v_i, v_j \rangle = \overline{v_j, v_i}$. Thus, $A_{ij} = \overline{A_{ji}}$.[28] In particular, $A_{ii} \in \mathbb{R}$.

Guiding Question
It is clear that the 2×2 matrix

$$A = \begin{pmatrix} 1 & i \\ -i & 0 \end{pmatrix}$$

satisfies the condition $A_{ij} = \overline{A_{ji}}$. However, can matrix A be the matrix representation of some $\langle \cdot, \cdot \rangle$?

Answer. *The* 0 *entry in* A *implies that* $\langle v_2, v_2 \rangle = 0$, *which occurs if and only if* $v_2 = 0$. *However,* v_2 *is a basis vector, so it must be nonzero. Therefore,* A *does not represent an inner product.*

This gives us another condition of matrix representations of inner products:

- By positivity, $\langle v_i, v_i \rangle \geq 0$. Since v_i is a basis vector, it follows that $\langle v_i, v_i \rangle > 0$. So, the elements on the diagonal of A must be positive real numbers.

However, even these two conditions are not enough to guarantee that matrix A represents an inner product! We will come back to these conditions in a later lecture.

Note that matrix A is a concrete representation of an inner product on V. For instance, you could input matrix A into a computer and be able to compute the inner product for any two vectors in V, without the computer having to understand anything about abstract inner product spaces.

16.3 Gram-Schmidt Procedure (continued)

In this section, we will discuss some consequences of the Gram-Schmidt Procedure.

Theorem 16.1
Suppose V is a finite-dimensional inner product space. Then,

(a) V has an orthonormal basis,

(b) if $e_1, \ldots, e_m$ is an orthonormal list, then it can be extended to an orthonormal basis.

[28] Matrices that satisfy this condition are known as Hermitian matrices. Note that if the entries of a matrix are all real, Hermitian matrices are the same as symmetric matrices.

Proof. For (a), let $v_1, \dots, v_n$ be an arbitrary basis of V. Then, applying Gram-Schmidt gives an orthonormal list $e_1, \dots, e_n$ which has length $\dim V$, so $e_1, \dots, e_n$ is an orthonormal basis

For (b), extend the orthonormal list to an arbitrary basis $e_1, \dots, e_m, v_{m+1}, \dots, v_n$ of V. Then, applying Gram-Schmidt gives an orthonormal list $e_1, \dots, e_m, e_{m+1}, \dots, e_n$, where the first m vectors are unchanged because they are already orthonormal. This orthonormal list has length $\dim V$, so it is an orthonormal basis. □

Now, consider the following statement.

Theorem 16.2
Suppose $e_1, \dots, e_m$ is the result of applying Gram-Schmidt on a linearly independent list of vectors $v_1, \dots, v_m$ in V. Then,

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

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

Proof. By Algorithm 15.11, we know that $e_i \in \operatorname{span}(v_1, \dots, v_i)$, which implies that

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

Since both of the above lists are linearly independent, both subspaces have dimension j, and thus are equal. □

Now, we will introduce a theorem which is a stronger version of Corollary 13.5.

Theorem 16.3 (Schur's Theorem)
Suppose V is an inner product space over $\mathbb{C}$ and $T \in \mathcal{L}(V)$. Then, there exists some orthonormal basis of V under which $\mathcal{M}(T)$ is upper triangular.

Proof. By Corollary 13.5, there exists some basis $v_1, \dots, v_n$ such that $\mathcal{M}(T)$ is upper-triangular. This implies that

$$T(v_i) \in \operatorname{span}(v_1, \dots, v_i).$$

Now, applying Gram-Schmidt to $v_1, \dots, v_n$ gives an orthonormal basis $e_1, \dots, e_n$. We will show that $\mathcal{M}(T, \{e_1, \dots, e_n\})$ is also upper-triangular. It is equivalent to show that

$$T(e_i) \in \operatorname{span}(e_1, \dots, e_i)$$

for all $i = 1, \dots, n$.

First, since $e_1 \in \operatorname{span}(v_1)$, it follows that $e_1 = cv_1$ for some scalar c. Also, since $T(v_1) \in \operatorname{span}(v_1)$, it follows that

$$\operatorname{span}(T(v_1)) \subset \operatorname{span}(\operatorname{span}(v_1)) = \operatorname{span}(v_1).$$

Then,

$$T(e_1) = cT(v_1) \in \operatorname{span}(T(v_1)) \subset \operatorname{span}(v_1) = \operatorname{span}(e_1).$$

For $T(e_2)$, by similar logic as above, we can show that

$$T(e_2) \in \operatorname{span}(T(v_1), T(v_2)) \subset \operatorname{span}(\operatorname{span}(v_1, v_2)) = \operatorname{span}(v_1, v_2) = \operatorname{span}(e_1, e_2),$$

where the last equality follows from Theorem 16.2. In the general case,

$$T(e_i) \in \operatorname{span}(T(v_1), \dots, T(v_i) \in \operatorname{span}(\operatorname{span}(v_1, \dots, v_i)) = \operatorname{span}(v_1, \dots, v_i) = \operatorname{span}(e_1, \dots, e_i),$$

as desired. □

16.4 Linear Functionals on Inner Product Spaces

Recall that an inner product $\langle \cdot, \cdot \rangle$ is not a function on V, but rather a function on two vectors in V. In terms of notation,

$$\langle \cdot, \cdot \rangle : V \times V \longrightarrow \mathbb{C}.$$

But, if we fix some vector $v \in V$, then the inner product becomes a function on one vector in V:

$$\begin{aligned} \langle \cdot, v \rangle : V &\longrightarrow \mathbb{C}, \\ u &\longmapsto \langle u, v \rangle. \end{aligned}$$

Then, by linearity in the first variable, it follows that $\langle \cdot, v \rangle$ is a linear functional on V. Recall the definition of linear functionals from Definition —.

It follows that the assignment

$$v \longmapsto \langle \cdot, v \rangle$$

gives a map from

$$V \longrightarrow \mathcal{L}(V, \mathbb{C}) = V',$$

where V' is the dual space of V (see Definition —).

Theorem 16.4 (Riesz Representation Theorem)
Suppose V is a finite-dimensional vector space over $\mathbb{C}$. Then, the map $T : V \to V'$ defined by

$$T(v) = \langle \cdot, v \rangle$$

is conjugate-linear and bijective.

Proof. First, we will check that T is conjugate-linear; that is, we will check that T satsifies the following two properties:

- T is additive: $T(v_1 + v_2) = T(v_1) + T(v_2)$,
- T is conjugate-homogeneous: $T(cv) = \overline{c}T(v)$.

Let $u \in V$. Then,

$$T(v_1 + v_2)(u) = \langle u, v_1 + v_2 \rangle = \langle u, v_1 \rangle + \langle u, v_2 \rangle = T(v_1)(u) + T(v_2)(u) = \big(T(v_1) + T(v_2)\big)(u),$$

so $T(v_1 + v_2) = T(v_1) + T(v_2)$. Furthermore,

$$T(cv)(u) = \langle u, cv \rangle = \overline{c}\langle u, v \rangle = \overline{c}\big(T(v)(u)\big) = \big(\overline{c}T(v)\big)(u),$$

so $T(cv) = \overline{c}T(v)$. Thus, T is conjugate-linear.

Next, we will check that T is injective. First, we will show that $T(v) = 0$ implies $v = 0$.[29] Let $v \in V$ such that $T(v) = 0$. Then,

$$\big(T(v)\big)(v) = \langle v, v \rangle = 0,$$

which implies that $v = 0$ by definiteness of the inner product. Now, let $v_1, v_2 \in V$ such that $T(v_1) = T(v_2)$. Then,

$$T(v_1 - v_2) = T(v_1) - T(v_2) = 0,$$

so $v_1 - v_2 = 0$. Therefore, T is injective.[30]

Now, we will show that T is surjective.[31] Let $e_1, \dots, e_n$ be an orthonormal basis of V. For any $\varphi \in V'$, let

$$S(\varphi) = \overline{\varphi(e_1)}e_1 + \overline{\varphi(e_2)}e_2 + \cdots + \overline{\varphi(e_n)}e_n.$$

[29] We have shown previously that if T is linear, then this condition would be su□cient to show T is injective. However, we only know that T is conjugate-linear, so we cannot directly apply this criterion.
[30] Note that even if T is conjugate-linear, the condition $T(v) = 0 \Rightarrow v = 0$ is su□cient to show that T is injective.
[31] Similarly, if T was linear, then injectivity would directly imply surjectivity because $\dim V = \dim V'$ (see —).

We will show that $T(S(\varphi)) = \varphi$, which implies that T is surjective. We know that

$$\begin{aligned} T(S(\varphi)) &= T(\overline{\varphi(e_1)}e_1 + \cdots + \overline{\varphi(e_n)}e_n) \\ &= \varphi(e_1)T(e_1) + \cdots + \varphi(e_n)T(e_n). \end{aligned}$$

Then, for any e_i,

$$\begin{aligned} T(S(\varphi))(e_i) &= \varphi(e_1)T(e_1)(e_i) + \cdots + \varphi(e_n)T(e_n)(e_i) \\ &= \varphi(e_1)\langle e_i, e_1\rangle + \cdots + \varphi(e_n)\langle e_i, e_n\rangle \\ &= \varphi(e_i)\langle e_i, e_i\rangle \\ &= \varphi(e_i). \end{aligned}$$

Since $T(S(\varphi))$ and φ agree on all basis vectors, it follows that $T(S(\varphi)) = \varphi$, as desired.[32] Therefore, T is both injective and surjective, so T is bijective. ☐

An consequence of this result is that given any linear functional φ, then $\varphi(u) = \langle u, v\rangle$ for

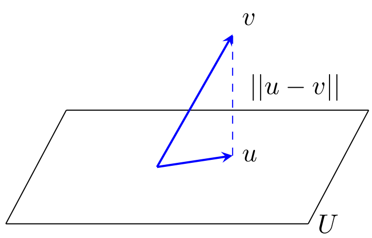
$$v = \overline{\varphi(e_1)}e_1 + \cdots + \overline{\varphi(e_n)}e_n$$

for any orthonormal basis $e_1, \ldots, e_n$. Furthermore, the Riesz Representation Theorem tells us that v is unique, so the right side of the above equation is the same regardless of which orthonormal basis $e_1, \ldots, e_n$ is chosen.

16.5 Orthogonal Projection

Now, we will shift our discussion to the notion of orthogonal projection. As we will see later, orthogonal projection actually has significant applications to minimization problems.

To visualize orthogonal projection, suppose we have a 2D plane that is a subspace of $\mathbb{R}^3$. Let v be a vector in $\mathbb{R}^3$. We wish to find a vector u such that u is in the plane and the distance between u and v is minimized, where $\text{dist}(u, v) = ||u - v||$.



Intuitively, the distance would be minimized when $u - v$ is perpendicular to the plane. This motivates the following definition.

Definition 16.5 (orthogonal complement, $U^\perp$)
Suppose U is a subspace of V. The **orthogonal complement** of U, denoted $U^\perp$, is defined by

$$U^\perp = \{v \in V | v \perp u \text{ for all } u \in U\}.$$

From this definition, we can deduce the following basic properties of $U^\perp$:

- $U^\perp$ is a subspace,
- $\{0\}^\perp = V$,
- $V^\perp = \{0\}$.

Furthermore, we also have the following lemma.

[32] While not needed for the proof, it is also true that $S(T(v)) = v$ for all $v \in V$, so S is the inverse of T.

Lemma 16.6
Suppose U is a finite-dimensional subspace of V. Then,

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

Proof. First, we will show that $U \cup U^{\perp} = \{0\}$. Let $u \in U \cup U^{\perp}$. It follows that u is orthogonal to itself, so $\langle u, u \rangle = 0$. By definiteness, it follows that $u = 0$.

Now, we will show that $V = U + U^{\perp}$. Let $e_1, \dots, e_m$ be an orthonormal basis of U. We wish to show that for any $v \in V$, there exists some $w \in U^{\perp}$ such that

$$v = c_1 e_1 + \cdots + c_m e_m + w.$$

Then, note that

$$\langle v, e_i \rangle = c_1 \langle e_1, e_i \rangle + \cdots + c_m \langle e_m, e_i \rangle + \langle w, e_i \rangle = c_i.$$

It follows that

$$w = v - \langle v, e_1 \rangle e_1 - \cdots - \langle v, e_m \rangle e_m.$$

It remains to check that $w \in U^{\perp}$, which is equivalent to showing that $\langle w, e_i \rangle$ for all $i = 1, \dots, m$. Then,

$$\begin{aligned} \langle w, e_i \rangle &= \big\langle v - \langle v, e_1 \rangle e_1 - \cdots - \langle v, e_m \rangle e_m, e_i \big\rangle \\ &= \langle v, e_i \rangle - \langle v, e_i \rangle \langle e_i, e_i \rangle \\ &= 0, \end{aligned}$$

as desired. Therefore, $V = U \oplus U^{\perp}$. □

Now, we can formally define orthogonal projection.

Definition 16.7 (orthogonal projection, P_U)
Suppose U is a finite-dimensional subspace of V. The **orthogonal projection** of V onto U is the linear map $P_U : V \to U$ defined by the following: for $v \in V$, write $v = u + w$, where $u \in U$ and $w \in U^{\perp}$. Then, $P_U(v) = u$.

A consequence of the proof of Lemma 16.6 is that if $e_1, \dots, e_m$ is an orthonormal basis of U, then

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

In the next lecture, we will discuss applications of orthogonal projections.

17 Orthogonal Projection (continued) and Self-Adjoint Operators

17.1 Review

Last time, we ended with a discussion on orthogonal projection. We defined the orthogonal complement (see Definition 16.5), denoted $U^\perp$, and proved that

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

We then defined the notion of orthogonal projection (see 16.7), denoted P_U, where $P_U(v)$ can be seen as taking the "U-part" of v. In particular, we showed that if given any orthonormal basis $e_1, \ldots, e_n$ of U, then

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

17.2 Applications of Orthogonal Projection

As we will see, P_U solves the following optimization problem.

Proposition 17.1
Suppose U is a finite-dimensional subspace of V and $v \in V$. Then, $P_U(v)$ has the shortest distance to v among all $u \in U$. In mathematical terms,

$$||v - u|| \geq ||v - P_U(v)||$$

for all $u \in U$ and equality occurs if and only if $u = P_U(v)$.

Proof. We can express v as

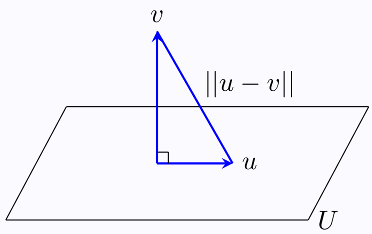
$$v = P_U(v) + w,$$

where $w \in U^\perp$. For any $u \in U$,

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

where the second equality follows from the Pythagorean Theorem (see Lemma 14.14), which applies because $P_U(v) - u \in U$ and $w \in U^\perp$. Since $w = v - P_U(v)$, it follows that $||v - u|| \geq ||v - P_U(v)||$, as desired. Equality holds when $||P_U(v) - u|| = 0$, which occurs if and only if $w = P_U(v)$. ☐

Example 17.2
Suppose U is a plane in $\mathbb{R}^3$.



Then, the shortest distance from v to any $u \in U$ is uniquely determined by setting u to be the orthogonal projection of v onto the plane. We can see that any other vector u would have a larger distance to v by drawing a right triangle, where $||v - u||$ is strictly larger than $||v - P_U(v)||$.

Orthogonal projections can also be used to create polynomial approximations of arbitrary functions.

Example 17.3
Suppose we wish to find a polynomial of degree at most 5 that approximates $\sin x$ as well as possible on the interval $[-\pi, \pi]$.[a]

Let f, g be two continuous functions on the interval $[-\pi, \pi]$. We define the distance between f and g as

$$\big(\text{dist}(f,g)\big)^2 = \int_{-\pi}^{\pi} |f(x) - g(x)|^2 \, dx.$$

[b] It follows that

$$||f||^2 = \int_{-\pi}^{\pi} |f(x)|^2 \, dx.$$

This definition of norm comes from the inner product

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x)\overline{g(x)} \, dx$$

over all continuous functions on $[-\pi, \pi]$.

Let $V = C([-\pi, \pi])$ denote the inner product space of all continuous functions on $[-\pi, \pi]$ with inner product defined as above and $U = \mathcal{P}_5(\mathbb{C})$. It is clear that U is a subspace of $C([-\pi, \pi])$. Now, our approximation problem can be reformulated to finding $u \in U$ such that $||v - u||$ is minimized, where $v = \sin(x)$.

By Proposition 17.1, the solution to this problem is $u = P_U(v)$. Apply the Gram-Schmidt Procedure on the basis $1, x, x^2, x^3, x^4, x^5$ of U to produce an orthonormal basis $e_1, e_2, e_3, e_4, e_5, e_6$ of U. Then, it follows that

$$P_U(v) = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_6 \rangle e_6.$$

Performing this computation results in

$$u = 0.987862x - 0.155271x^3 + 0.00564312x^5,$$

which is shown in the textbook to be a much more accurate approximation of $\sin x$ than the Taylor polynomial.

[a] Note that we must restrict ourselves to the interval $[-\pi, \pi]$ because a polynomial must go to $\pm\infty$ at the extremes, which would give a bad approximation of $\sin x$.
[b] This definition of distance is known as L^2-distance.

Note that $C([-\pi, \pi])$ in the above example is an infinite-dimensional vector space. However, we are projecting onto a finite-dimensional vector space U, so all of our theorems still hold.

17.3 Adjoints

We will now shift our discussion to adjoints of linear maps.

Definition 17.4 (adjoint)
Suppose V, W are inner product spaces and $T : V \to W$ is a linear map. Then, the **adjoint** of T is the map $T^* : W \to V$ such that

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

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

Note that the above definition doesn't directly tell you what T^*w is for any $w \in W$. Instead, it tells you that $\langle v, T^*w \rangle$ is determined by T for all $v \in V$, which in turn determines T^*w.

To see why this is the case, let $e_1, \ldots, e_n$ and $f_1, \ldots, f_m$ be orthonormal bases of V and W, respectively. Let $\mathcal{M}(T, (e_1), (f_j)) = (A_{ji})$, so

$$T(e_i) = A_{1i}f_1 + A_{2i}f_2 + \cdots + A_{mi}f_m.$$

It follows that

$$\begin{aligned}\langle e_i, T^* f_j\rangle &= \langle Te_i, f_j\rangle \\ &= \langle A_{1i} f_1 + A_{2i} f_2 + \cdots + A_{mi} f_m, f_j\rangle \\ &= A_{ji},\end{aligned}$$

which implies that

$$T^* f_j = \overline{A_{j1}} e_1 + \overline{A_{j2}} e_2 + \cdots + \overline{A_{jn}} e_n.$$

The same logic as above can be expanded to show that $T^* w$ is uniquely determined for any vector $w = c_1 f_1 + \cdots + c_m f_m \in W$. Thus, T^* is well-defined.

Theorem 17.5
T^* is a linear map.

Proof. First, we will show that T^* is additive. We wish to show that $T^*(w_1 + w_2) - T^*(w_1) - T^*(w_2) = 0$ for arbitrary $w_1, w_2 \in W$. For any $v \in V$,

$$\begin{aligned}\langle v, T^*(w_1 + w_2) - T^*(w_1) - T^*(w_2)\rangle &= \langle v, T^*(w_1 + w_2)\rangle - \langle v, T^*(w_1)\rangle - \langle v, T^*(w_2)\rangle \\ &= \langle Tv, w_1 + w_2\rangle - \langle Tv, w_1\rangle - \langle Tv, w_2\rangle \\ &= 0.\end{aligned}$$

Since v is arbitrary, it follows that $T^*(w_1 + w_2) - T^*(w_1) - T^*(w_2) = 0$. Therefore, T^* is additive.

Next, we will show that T^* is homogeneous. We wish to show that $T^*(cw) - cT^*(w)$. Then, using a similar process to above,

$$\begin{aligned}\langle v, T^*(cw) - cT^*(w)\rangle &= \langle v, T^*(cw)\rangle - \langle v, cT^*(w)\rangle \\ &= \langle Tv, cw\rangle - \bar{c}\langle v, T^* w\rangle \\ &= \bar{c}\langle Tv, w\rangle - \bar{c}\langle Tv, w\rangle \\ &= 0,\end{aligned}$$

so $T^*(cw) - cT^*(w) = 0$. Therefore, T^* is homogeneous, which implies that T^* is linear. □

Since T^* is a linear map, we can represent T^* with a matrix. Furthermore, we showed earlier that $T^* f_j = \overline{A_{j1}} e_1 + \cdots + \overline{A_{jn}} e_n$, so

$$\mathcal{M}(T^*, (f_j), (e_i)) = (\overline{A_{ij}}).$$

Therefore, $\mathcal{M}(T^*)$ is the ***conjugate transpose*** of $\mathcal{M}(T)$, which is obtained by taking the transpose and taking the conjugate of each entry.

Example 17.6
Consider the identity map. It follows that

$$\langle \mathrm{id}(v), w\rangle = \langle v, w\rangle = \langle v, \mathrm{id}^*(w)\rangle,$$

so $\mathrm{id}^* = \mathrm{id}$.

More generally, consider the scalar operator $c \cdot \mathrm{id}$. Then,

$$\langle (c \cdot \mathrm{id})(v), w\rangle = c\langle v, w\rangle = \langle v, \bar{c} w\rangle = \langle v, (c \cdot \mathrm{id})^*(w)\rangle,$$

so $(c \cdot \mathrm{id})^* = \bar{c} \cdot \mathrm{id}$.

Additionally, these are some basic properties of adjoints.

Fact 17.7 (Properties of adjoints)

1. $(T_1 + T_2)^* = T_1^* + T_2^*$
2. $(cT)^* = \overline{c}T^*$
3. $(TS)^* = S^*T^*$

17.4 Self-Adjoint Operators

Now, we will shift our discussion to operators on inner product spaces. First, we will discuss the notion of self-adjoint operators.

Definition 17.8 (self-adjoint)
An operator $T \in \mathcal{L}(V)$ is **self-adjoint** if $T = T^*$. In other words, T is self-adjoint if and only if

$$\langle Tv, w\rangle = \langle v, Tw\rangle$$

for all $v, w \in V$.

Suppose $e_1, \ldots, e_n$ is an orthonormal basis of complex vector space V. Since $\mathcal{M}(T^*)$ is the conjugate transpose of $\mathcal{M}(T)$ under any orthonormal basis, it follows that T is self-adjoint if and only if $\mathcal{M}(T)$ is invariant under conjugate transpose.[33] These matrices take the form

$$\begin{pmatrix} A_{11} & & & A_{ij} \\ & A_{22} & & \\ & & \ddots & \\ A_{ji} & & & A_{nn} \end{pmatrix},$$

where $A_{ij} = \overline{A_{ji}}$. In particular, this implies that $A_{ii} = \overline{A_{ii}}$, so the entries on the diagonal must be real.

If V is a real vector space, then T is self-adjoint if and only if $\mathcal{M}(T)$ is a symmetric matrix.

By definition, every eigenvalue is real over real vector spaces. Thus, the next result is only interesting over complex vector spaces.

Proposition 17.9
All eigenvalues of a self-adjoint operator are real.

Proof. Let λ be an eigenvalue of T with corresponding eigenvector v. Then,

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

Also,

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

It follows that $\lambda||v||^2 = \overline{\lambda}||v||^2$, which implies that $\lambda = \overline{\lambda}$ since v is nonzero. Thus, λ is real. □

Now, we will discuss an interesting way to think about self-adjoint operators. For any $T \in \mathcal{L}(V)$, we can consider the function φ defined by

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

Suppose we wish φ satisfied the property an inner product; for instance, it is easy to show that φ is linear in the first variable and conjugate-linear in the second variable.

When does φ satisfy conjugate symmetry? We can see that

$$\varphi(v, u) = \langle Tv, u\rangle = \overline{\langle u, Tv\rangle}.$$

[33] Such matrices are known as Hermitian matrices, which were introduced in the previous lecture.

It follows that

$$\varphi(u,v) = \langle Tv, u\rangle \quad \text{and} \quad \overline{\varphi(v,u)} = \langle u, Tv\rangle,$$

so φ satisfies conjugate symmetry if and only if T is self-adjoint.

Even with the condition that T is self-adjoint, note that φ is still not necessarily an inner product, since it may not satisfy positivity or definiteness.[34]

We will use the above definition of the function φ for our next result.

Proposition 17.10
Suppose $T \in \mathcal{L}(V)$ is self-adjoint. If $\langle Tv, v\rangle = 0$ for all $v \in V$, then $T = 0$.

Proof. Let $\varphi(u,v) = \langle Tu, v\rangle$. We know that

- φ is linear in the first variable and conjugate-linear in the second variable,
- φ has conjugate-symmetry,
- $\varphi(v,v) = 0$.

Then, it follows that

$$\begin{aligned}
\varphi(u+v, u+v) &= \varphi(u,u) + \varphi(u,v) + \varphi(v,u) + \varphi(v,v) \\
&= \varphi(u,v) + \varphi(v,u) \\
&= \varphi(u,v) + \overline{\varphi(u,v)} \\
&= 2\operatorname{Re}(\varphi(u,v)) \\
&= 0,
\end{aligned}$$

so $\operatorname{Re}(\varphi(u,v)) = 0$ for all $u, v \in V$. Similarly,

$$\begin{aligned}
\varphi(u+iv, u+iv) &= \varphi(u,u) + \varphi(u,iv) + \varphi(iv,u) + \varphi(v,iv) \\
&= \varphi(u,iv) + \varphi(iv,u) \\
&= i(\varphi(v,u) - \varphi(u,v)) \\
&= i(\overline{\varphi(u,v)} - \varphi(u,v)) \\
&= 2\operatorname{Im}(\varphi(u,v)) \\
&= 0,
\end{aligned}$$

so $\operatorname{Im}(\varphi(u,v)) = 0$ for all $u, v \in V$. Thus, $\varphi = 0$.

Now, for any $u \in V$,

$$\varphi(u, Tu) = \langle Tu, Tu\rangle = 0,$$

which implies $Tu = 0$ by definiteness. Therefore, $T = 0$. □

[34] Later, we will introduce the notion of positive operators. If T is a positive operator, then would make φ satisfy positivity.

18 The Spectral Theorem

18.1 Review

Last time, we introduced adjoints and ended with a discussion on self-adjoint operators (see Definition 17.8). We also proved some important properties of self-adjoint operators, in particular that all eigenvalues of a self-adjoint operator are real (see Proposition 17.9).

18.2 Self-Adjoint Operators (continued)

First, we will discuss some consequences of Proposition 17.9.

Example 18.1
Suppose $V = \mathbb{C}^n$ and $e_1, \ldots, e_n$ is the standard basis. Consider $T \in \mathcal{L}(V)$ such that

$$\mathcal{M}(T) = \begin{pmatrix} a_1 & & & 0 \\ & a_2 & & \\ & & \ddots & \\ 0 & & & a_n \end{pmatrix}.$$

Thus, $Te_i = a_i e_i$ for all $i = 1, 2, \ldots, n$.

Guiding Question
IF T is self-adjoint, then what are the possible values of $a_1, a_2, \ldots, a_n$?

Answer. *It is clear that $a_1, \ldots, a_n$ are eigenvalues of T. Thus, for T to be self-adjoint, $a_1, \ldots, a_n$ must all be real.*

Now, we will prove the converse of the above statement.

Proposition 18.2
Suppose V and $T \in \mathcal{L}(V)$ is defined as in Example 18.1. If all a_i are real, then T is self-adjoint.

Proof. Let $e_1, \ldots, e_n$ be the standard basis of $V = \mathbb{C}^n$. First, we will show that $\langle Te_i, e_j \rangle = \langle e_i, Te_j \rangle$. We find that

$$\langle Te_i, e_j \rangle = \langle a_i e_i, e_j \rangle = \begin{cases} 0 & i \neq j, \\ a_i & i = j. \end{cases}$$

and

$$\langle e_i, Te_j \rangle = \langle e_i, a_j e_j \rangle = \begin{cases} 0 & i \neq j, \\ a_j & i = j. \end{cases}$$

Since $a_i = a_j$ for $i = j$, it follows that $\langle Te_i, e_j \rangle = \langle e_i, Te_j \rangle$ for all pairs of basis vectors e_i, e_j. Then, by linearity in the first variable and conjugate linearity in the second variable, it follows that $\langle Tu, v \rangle = \langle u, Tv \rangle$ for any $u, v \in V$. Thus, T is self-adjoint. □

The next result is a more general statement of Proposition 18.2.

Proposition 18.3
Suppose V is a complex inner product space, $e_1, \ldots, e_n$ is an orthonormal basis of V, and $T \in \mathcal{L}(V)$ such that

$$\mathcal{M}(T, (e_1, \ldots, e_n)) = \begin{pmatrix} a_1 & & & 0 \\ & a_2 & & \\ & & \ddots & \\ 0 & & & a_n \end{pmatrix}$$

for $a_1, \ldots, a_n \in \mathbb{R}$. Then, T is self-adjoint.

Proof. This proof is identical to the proof of Proposition 18.2. ☐

Later in this lecture, we will prove the converse of Proposition 18.3; that is, if T is self-adjoint, then there exists an orthonormal basis such that T is diagonal with real entries.

18.3 Real Spectral Theorem

We will first prove a simpler version of the above statement over real vectors spaces. To do this, we will need the following two lemmas.

Note that by Theorem 13.1, any $T \in \mathcal{L}(V)$ for a complex vector space V has an eigenvalue. Thus, the next result is only interesting for real vector spaces.

Lemma 18.4
Suppose V is a $\mathbb{R}$-inner product space and $T \in \mathcal{L}(V)$ is self-adjoint. Then, T has an eigenvalue.

Proof. We wish to find some $\lambda \in \mathbb{R}$ such that $\text{Null}(T - \lambda \text{id}_V) \neq \{0\}$. Let $e_1, \ldots, e_n$ be an orthonormal basis of V and let $\mathcal{M}(T) = A$ with respect to this basis. It is equivalent to find some λ such that $\text{rank}(A - \lambda I) < n$.

By the definition of $\mathcal{M}(T)$, we know that A_{ij} is the coefficient of the e_i term of Te_j; in other words, $A_{ij} = \langle Te_j, e_i \rangle$. Then,

$$A_{ij} = \langle Te_j, e_i \rangle = \langle e_j, Te_i \rangle = \langle Te_i, e_j \rangle = Aji,$$

so A is symmetric. Since A is also a real matrix, it follows that A is a Hermitian matrix.

Now, view A as the matrix of an operator $S \in \mathcal{L}(\mathbb{C}^n)$. Since A is a Hermitian matrix, it follows that A is equivalent to its conjugate transpose, so S is self-adjoint. Then, by Theorem 13.1, S must have some eigenvalue λ. Since S is self-adjoint, it follows that λ is real.

Thus, since $\lambda \in \mathbb{R}$ is an eigenvalue of S, it follows that

$$\begin{aligned} &\text{Null}(S - \lambda \text{id}_{\mathbb{C}^n}) \neq \{0\} \\ \implies &\text{rank}(A - \lambda I) < n \\ \implies &\text{Null}(T - \lambda \text{id}_V) \neq \{0\}, \end{aligned}$$

so λ is an eigenvalue of T. ☐

Although this lemma concerns an operator T in a real vector space, the proof relies on creating an analogous operator S in some complex vector space such that $\mathcal{M}(T) = \mathcal{M}(S) = A$. This technique is known as ***complexification***, which we will not cover in detail in this course.

Lemma 18.5
Suppose $T \in \mathcal{L}(V)$ is self-adjoint and U is a T-invariant subspace of V. Then,

1. $U^\perp$ is T-invariant,
2. $T|_U$ is self-adjoint,
3. $T|_{U^\perp}$ is self-adjoint.

Proof. To prove (1), let $v \in U^\perp$ and $u \in U$. Then,

$$\langle Tv, u \rangle = \langle v, Tu \rangle = 0,$$

where the second equality holds because $v \in U^\perp$ and U is T-invariant, so $Tu \in U$. Because u is an arbitrary vector in U, it follows that $Tv \in U^\perp$, so $U^\perp$ is T-invariant.

To prove (2), for any $u, v \in U$,

$$\langle (T_U u, v \rangle = \langle Tu, v \rangle = \langle u, Tv \rangle = \langle u, T_U v \rangle,$$

so T_U is self-adjoint.

To prove (3), replace U with $U^\perp$ in the proof of (2). ☐

Now, we are ready to prove the Real Spectral Theorem.

Theorem 18.6 (Real Spectral Theorem)
Suppose V is a finite-dimensional $\mathbb{R}$-inner product space and $T \in \mathcal{L}(V)$. Then, the following are equivalent:

1. T is self-adjoint,
2. T has a diagonal matrix under some orthonormal basis of V.

Proof. By Proposition 18.3, (2) implies (1).

Now, we will show that (1) implies (2) by induction on $\dim V$. Let $n = \dim V$. If $n = 1$, then T has a diagonal matrix since every 1×1 matrix is diagonal.

Now, assume $n > 1$. By Lemma 18.4, T has an eigenvalue λ_1. Let e_1 be a corresponding eigenvector of λ_1 such that $||e_1|| = 1$ (which can be done by choosing any eigenvector and dividing it by its norm). Then, let $U = \operatorname{span}(e_1)$. It is clear that U is T-invariant. By Lemma 18.5, $U^\perp$ is T-invariant with dimension $n-1$ and $T|_{U^\perp}$ is self-adjoint.

By the inductive hypothesis, $T|_{U^\perp}$ has a diagonal matrix for some orthonormal basis $e_2, \ldots, e_n$ of $U^\perp$. In other words,

$$\begin{aligned} Te_2 &= \lambda_2 e_2 \\ &\vdots \\ Te_n &= \lambda_n e_n. \end{aligned}$$

Additionally, $Te_1 = \lambda_1 e_1$. Therefore, T is diagonal with respect to the orthonormal basis $e_1, \ldots, e_n$ of V. □

18.4 Normal Operators and Complex Spectral Theorem

Now, we will extend the Spectral Theorem to complex vector spaces.

Theorem 18.7 (Complex Spectral Theorem for self-adjoint operators)
Suppose V is a finite-dimensional $\mathbb{C}$-inner product space and $T \in \mathcal{L}(V)$. Then, the following are equivalent:

1. T is self-adjoint,
2. T has a diagonal matrix with real entries under some orthonormal basis of V.

Proof. The proof is identical to the proof of the Real Spectral Theorem, except we can apply Theorem 13.1 instead of Lemma 18.4. □

Now, suppose V is a complex vector space and $T \in \mathcal{L}(V)$ has a diagonal matrix with diagonal entries $a_1, \ldots, a_n$ for some orthonormal basis $e_1, \ldots, e_n$ of V, except $a_1, \ldots, a_n$ are *not necessarily real*. Then, since $\mathcal{M}(T^*)$ is the conjugate transpose of $\mathcal{M}(T)$, it follows that

$$Te_i = a_i e_i \quad \text{and} \quad T^* e_i = \overline{a_i} e_i.$$

In particular,

$$TT^* e_i = a_i \overline{a_i} e_i = T^* T e_i,$$

so $TT^* = T^*T$.

This motivates the following definition.

Definition 18.8 (normal)
An operator $T \in \mathcal{L}(V)$ is **normal** if

$$TT^* = T^*T.$$

Consider the following examples of normal operators.

Example 18.9
Suppose $T \in \mathcal{L}(\mathbb{C}^2)$ such that

$$\mathcal{M}(T) = \begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix}$$

with respect to the standard basis. It follows that

$$\mathcal{M}(T^*) = \begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}.$$

Since $\mathcal{M}(T) \neq \mathcal{M}(T^*)$, it follows that T is not self-adjoint. However, we can compute

$$\mathcal{M}(T)\mathcal{M}(T^*) = \begin{pmatrix} 13 & 0 \\ 0 & 13 \end{pmatrix} = \mathcal{M}(T^*)\mathcal{M}(T).$$

Thus, $TT^* = T^*T$, so T is normal.

Example 18.10
Suppose $T \in \mathcal{L}(V)$ is self-adjoint. Let $c \in \mathbb{C}$. We will show that cT is normal.

We know that

$$(cT)^* = \overline{c}T^* = \overline{c}T.$$

Then,

$$(cT)(cT)^* = cT \cdot \overline{c}T = |c|^2T^2 = \overline{c}T \cdot cT = (cT)^*(cT).$$

Thus, cT is normal.

Additionally, cT is self-adjoint if and only if $cT = \overline{c}T$, which occurs when c is real.

Before we prove the full statement of the Complex Spectral Theorem, we will need the following two lemmas.

Lemma 18.11
Suppose V is a $\mathbb{C}$-inner product space and $T \in \mathcal{L}(V)$ is normal. Then, there exists some nonzero $v \in V$ such that v is an eigenvector for both T and T^*.

Proof. By Theorem 13.1, there exists some λ which is an eigenvalue of T. We will show that $E(\lambda, T)$ is T^*-invariant.

Let $v \in E(\lambda, T)$. It follows that $Tv = \lambda v$. Then,

$$T(T^*v) = T^*(Tv) = T^*(\lambda v) = \lambda(T^*v),$$

where the first equality holds because T is normal. Thus, $T^*v \in E(\lambda, T)$, so $E(\lambda, T)$ is T^*-invariant.

Now, choose any eigenvector v of $T^*|_{E(\lambda,T)}$, which must exist by Theorem 13.1. It follows that $v \in E(\lambda, T)$, so v is also an eigenvector of T.[35] □

Lemma 18.12
Suppose V is a $\mathbb{C}$-inner product space, $T \in \mathcal{L}(V)$ and U is a T-invariant subspace of V. Then, $U^\perp$ is T^*-invariant.

Proof. Suppose $u \in U$ $v \in U^\perp$. Then,

$$\langle u, T^*v \rangle = \langle Tu, v \rangle = 0,$$

since $Tu \in U$ and $v \in U^\perp$. Thus, $T^*v \in U^\perp$, so $U^\perp$ is T^*-invariant. □

[35] It is actually the case that if v has corresponding eigenvalue λ for T, then v has corresponding eigenvalue $\overline{\lambda}$ for T^*. This property is not needed for our discussion, so we did not prove it here.

Now, we are ready to prove the full statement of the Complex Spectral Theorem.

Theorem 18.13 (Complex Spectral Theorem for normal operators)
Suppose V is a finite-dimensional $\mathbb{C}$-inner product space and $T \in \mathcal{L}(V)$. Then, the following are equivalent:

1. T is normal,
2. T has a diagonal matrix under some orthonormal basis of V.

Proof. First, we will show that (2) implies (1). Since T^* is the conjugate transpose of T, it follows that T^* is also diagonal. Any two diagonal matrices commute, so T commutes with T^*. Thus, T is normal.

Now, we will show that (1) implies (2) by induction on $\dim V$. Let $n = \dim V$. If $n = 1$, then T has a diagonal matrix since every 1×1 matrix is diagonal.

Then, assume $n > 1$. By Lemma 18.11, there exists some vector e_1 such that $||e_1|| = 1$ and is an eigenvector for both T and T^*. Let $U = \text{span}(e_1)$. It follows that U is both T-invariant and T^*-invariant. By Lemma 18.12, $U^\perp$ is T^*-invariant. Furthermore, since U is also T^*-invariant, it follows that $U^\perp$ is also T-invariant because $(T^*)^* = T$. Additionally,

$$(T|_{U^\perp})(T|_{U^\perp})^* = T|_{U^\perp}(T^*)|_{U^\perp} = (T^*)|_{U^\perp}T|_{U^\perp} = (T|_{U^\perp})^*(T|_{U^\perp}),$$

so $T|_{U^\perp}$ is normal.

By the inductive hypothesis, there exists an orthonormal basis $e_2, \ldots, e_n$ of $U^\perp$ such that the matrix of $T|_{U^\perp}$ is diagonal. Thus, $e_1, \ldots, e_n$ is an orthonormal basis of V such that T is diagonal. $\square$

The above result is quite fascinating: the definition of normal operators doesn't say anything about the eigenvalues or matrix of T. But, from the simple fact that T commutes with T^*, we can deduce that T is diagonalizable.

18.5 Normal Operators (continued)

Now, we will use the Complex Spectral Theorem to deduce some properties of normal operators.

Theorem 18.14 (Properties of normal operators)
Suppose $T \in \mathcal{L}(V)$ is normal. Then,

1. $||Tv|| = ||T^*v||$ for all $v \in V$,
2. $E(\lambda, T) = E(\overline{\lambda}, T^*)$ for all $\lambda \in \mathbb{C}$,
3. $E(\lambda, T) \perp E(\mu, T)$ for $\lambda, \mu \in \mathbb{C}$ and $\lambda \neq \mu$.

Proof. To prove (1), by the Spectral Theorem, there exists some orthonormal basis $e_1, \ldots, e_n$ such that T is diagonal. It follows that $Te_i = \lambda_i e_i$ and $T^*e_i = \overline{\lambda_i} e_i$ because T^* is the conjugate tranpose of T. Let $v = a_1e_1 + \cdots + a_ne_n \in V$. Then,

$$\begin{aligned} ||Tv||^2 &= ||a_1\lambda_1e_1 + \cdots + a_n\lambda_ne_n|| \\ &= |a_1\lambda_1|^2 + \cdots |a_n\lambda_n|^2 \\ &= |a_1\overline{\lambda_1}|^2 + \cdots |a_n\overline{\lambda_n}|^2 \\ &= ||a_1\overline{\lambda_1}e_1 + \cdots + a_n\overline{\lambda_n}e_n|| \\ &= ||T^*v||^2, \end{aligned}$$

as desired.[36]

To prove (2), consider an orthonormal basis $e_1, \ldots, e_n$ such that T is diagonal. It follows that

$$E(\lambda, T) = \text{span}(e_i | \lambda_i = \lambda).$$

[36] This statement is actually an if and only if statement; that is, T is normal if and only if $||Tv|| = ||T^*v||$ for all $v \in V$.

Since T^* is the conjugate transpose of T, it follows that

$$E(\overline{\lambda}, T) = \text{span}(e_i | \overline{\lambda_i} = \bar{l}) = \text{span}(e_i | \lambda_i = \lambda),$$

so $E(\lambda, T) = E(\overline{\lambda}, T^*)$.

To prove (3), we know that

$$E(\lambda, T) = \text{span}(e_i | \lambda_i = \lambda) \quad \text{and} \quad E(\mu, T) = \text{span}(e_j | \lambda_j = \mu).$$

Because $\lambda \neq \mu$, there is no e_i that is included in both $E(\lambda, T)$ and $E(\mu, T)$. Finally, since $e_1, \ldots, e_n$ is an orthonormal list, it follows that $E(\lambda, T) \perp E(\mu, T)$. □

19 Isometries and Singular Value Decomposition

19.1 Review

Last time, we discussed the Real and Complex Spectral Theorems (recall Theorem 18.6 and Theorem 18.13).

19.2 Isometries

In this section, we will consider linear maps that preserve norms.

Definition 19.1 (isometry)
Suppose V and W are inner product spaces. A linear map $S : V \to W$ is an *isometric embedding* if

$$||Sv|| = ||v||$$

for all $v \in V$. If S is also surjective, then S is an **isometry**.

Fact 19.2
Suppose S is an isometric embedding, Then, S is injective.

Proof. A linear map is injective if $Sv = 0$ implies $v = 0$. Let $v \in V$ such that $Sv = 0$. Then,

$$||v|| = ||Sv|| = 0,$$

so $v = 0$. Therefore, S is injective. □

Thus, an isometry S is an isomorphism because it is both injective and surjective. In particular, an isometric embedding $S \in \mathcal{L}(V, W)$ is an isometry when V and W have the same dimension.

Example 19.3

Suppose $V = \mathbb{R}^2$ with the standard inner product and $W = \mathbb{R}^2$ with inner product defined by

$$\langle e_1', e_1' \rangle = 2,$$
$$\langle e_1', e_2' \rangle = 0,$$
$$\langle e_2', e_2' \rangle = \frac{1}{2},$$

for a basis e_1', e_2' of $\mathbb{R}^2$.

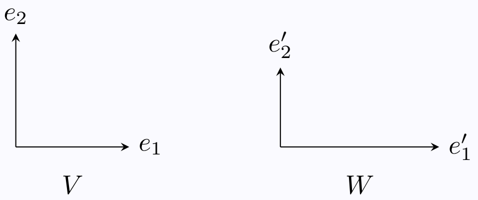
It follows that

$$||(x, y)||_V = \sqrt{x^2 + y^2}$$

and

$$\begin{aligned} ||(x, y)||_W &= \sqrt{\langle xe_1' + ye_2', xe_1' + ye_2' \rangle} \\ &= \sqrt{x^2\langle e_1', e_1' \rangle + 2xy\langle e_1', e_2' \rangle + y^2\langle e_2', e_2' \rangle} \\ &= \sqrt{2x^2 + \frac{1}{2}y^2}. \end{aligned}$$

Now, we wish to find an isometry $S : V \to W$. Geometrically, we can visualize the following diagram for V and W:



It is clear that setting $Se_1 = e_1'$ would not be an isometry because $||e_1||_V = 1$ and $||e_1'||_W = \sqrt{2}$. However, this implies that setting $Se_1 = \frac{1}{\sqrt{2}}e_1'$ does conserve norms. By similar logic, setting $Se_2 = \sqrt{2}e_2'$ also conserves norms. Furthermore, $\langle Se_1, Se_2 \rangle = \langle \frac{1}{\sqrt{2}e_1'}, \sqrt{2}e_2' \rangle = 0$. Then, for any $v = xe_1 + ye_2 \in V$,

$$\begin{aligned} ||Sv||_W &= ||\tfrac{1}{\sqrt{2}}xe_1' + \sqrt{2}ye_2'|| \\ &= \sqrt{2(\tfrac{1}{\sqrt{2}}x)^2 + \tfrac{1}{2}(\sqrt{2}y)^2} \\ &= \sqrt{x^2 + y^2} \\ &= ||v||_V. \end{aligned}$$

Therefore, S is an isometry.

This example is one out of many possible isometries from V to W. For instance, it is easy to verify by the same logic as above that the linear map defined by

$$Se_1 = \sqrt{2}e_2' \quad \text{and} \quad Se_2 = \tfrac{1}{\sqrt{2}}e_1'$$

is an isometry.

Now, we will discuss some properties of isometries.

Theorem 19.4 (Properties of isometries)
Suppose $S \in \mathcal{L}(V, W)$ is an isometry. Then,

1. $\langle Su, Sv \rangle = \langle u, v \rangle$ for all $u, v \in V$,
2. $S^*S = \mathrm{id}_V$
3. if $e_1, \ldots, e_n$ is an orthonormal basis of V, then $Se_1, \ldots, Se_n$ is an orthonormal list of W.

Additionally, for arbitrary $S \in \mathcal{L}(V, W)$,

4. if $e_1, \ldots, e_n$ is an orthonormal basis of V and $Se_1, \ldots, Se_n$ is an orthonormal list of W, then S is an isometry.

Proof. To prove (1), we know that $||S(u + v)|| = ||u + v||$. Then, we can compute

$$\begin{aligned} ||S(u+v)||^2 &= \langle Su + Sv, Su + Sv \rangle \\ &= \langle Su, Su \rangle + \langle Su, Sv \rangle + \langle Sv, Su \rangle + \langle Sv, Sv \rangle \\ &= ||Su||^2 + ||Sv||^2 + 2\,\mathrm{Re}\langle Su, Sv \rangle. \end{aligned}$$

Similarly, we compute

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

It follows that

$$||Su||^2 + ||Sv||^2 + 2\,\mathrm{Re}\langle Su, Sv \rangle = ||u||^2 + ||v||^2 + 2\,\mathrm{Re}\langle u, v \rangle.$$

Since S is an isometry, we know that $||Su|| = ||u||$ and $||Sv|| = ||v||$. Therefore, $\mathrm{Re}\langle Su, Sv \rangle = \mathrm{Re}\langle u, v \rangle$ for all $u, v \in V$.

Then, replacing u with iu in the above argument gives us

$$\mathrm{Re}(i\langle Su, Sv \rangle = \mathrm{Re}(i\langle u, v \rangle),$$

which implies that $\mathrm{Im}\langle Su, Sv \rangle = \mathrm{Im}\langle u, v \rangle$. Thus, $\langle Su, Sv \rangle = \langle u, v \rangle$ for all $u, v \in V$.

To prove (2), we know that

$$\langle u, S^*Sv \rangle = \langle Su, Sv \rangle = \langle u, v \rangle,$$

where the second equality follows from (1). It follows that $\langle u, S^*S - v \rangle = 0$ for all $u, v \in V$. This implies that $S^*S - v$ is orthogonal to all vectors in V; in particular, it must be orthogonal to itself, so $S^*S - v = 0$. Therefore, $S^*S = \mathrm{id}_V$.

To prove (3), we know that $\langle Se_i, Se_j \rangle = \langle e_i, e_j \rangle$ by (1). Since $e_1, \ldots, e_n$ is orthonormal, it follows that

$$\langle Se_i, Se_j \rangle = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases}$$

so $Se_1, \ldots, Se_n$ is orthonormal.

To prove (4), suppose $v = x_1 e_1 + \cdots + x_n e_n \in V$. Then,

$$\begin{aligned} ||v||^2 &= ||x_1 e_1||^2 + \cdots + ||x_n e_n||^2 \\ &= |x_1|^2 + \cdots |x_n|^2 \end{aligned}$$

and

$$\begin{aligned} ||Sv||^2 &= ||x_1 Se_1 + \cdots x_n Se_n||^2 \\ &= ||x_1 Se_1||^2 + \cdots + ||x_n Se_n||^2 \\ &= |x_1|^2 + \cdots |x_n|^2. \end{aligned}$$

Thus, $||Sv|| = ||v||$, so S is an isometry. □

Properties (3) and (4) imply that $S \in \mathcal{L}(V, W)$ sending an orthonormal basis of V to an orthonormal list of W is equivalent to S being an isometry. In fact, the first two properties are also equivalent to S being an isometry, the details of which are in the textbook.

19.3 Singular Values

Now, let $S \in \mathcal{L}(V, W)$ be any linear map such that $\dim V \leq \dim W$.

Guiding Question
Can we find orthonormal bases $e_1, \ldots, e_n$ of V and $f_1, \ldots, f_m$ of W such that

$$S(e_i) = a_i f_i$$

for $i = 1, 2, \ldots, n$?

In other words, we want

$$\mathcal{M}(S, (e_i), (f_j)) = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \\ & & \\ 0 & \cdots & 0 \end{pmatrix}.$$

We will spend this section showing that finding such orthonormal is always possible.

Note that $a_1, \ldots, a_n$ are not eigenvalues of S, since $e_1, \ldots, e_n$ and $f_1, \ldots, f_m$ are different bases. Instead, we call $a_1, \ldots, a_n$ *singular values* of S. We will show that singular values only depend on S; namely, they are independent of the chosen orthonormal bases of V and W.

First, we will further explore $a_1, \ldots, a_n$. By definition, $S(e_i) = a_i f_i$. Then, the matrix of $S^* \in \mathcal{L}(W, V)$ is the conjugate transpose of S,

$$\mathcal{M}(S^*, (f_j), (e_i)) = \begin{pmatrix} \overline{a_1} & & & 0 \\ & \ddots & & \vdots \\ & & \overline{a_n} & 0 \end{pmatrix}.$$

Thus,

$$S^*(f_j) = \begin{cases} \overline{a_j} e_j & \text{for } 1 \leq j \leq n, \\ 0 & \text{for } j > n. \end{cases}$$

Now, note that $S^*S \in \mathcal{L}(V)$. Then,

$$S^*S(e_i) = S^*(\overline{a_i} f_i) = a_i \overline{a_i} e_i = |a_i|^2 e_i.$$

Therefore, $|a_i|^2$ are the eigenvalues of S^*S with corresponding eigenvectors e_i. Since the eigenvalues of S^*S is independent of the chosen basis of V, it follows that $|a_i|$ is intrinsic to S. In conclusion, the singular values $|a_1|, \ldots, |a_n|$ are the square roots of the eigenvalues of S^*S.[37]

Definition 19.5 (singular values)
Suppose $S \in \mathcal{L}(V)$. The **singular values** of S are the square roots of the eigenvalues of S^*S.

Note that we reached the above definition by assuming that there exist appropriate orthonormal bases $e_1, \ldots, e_n$ and $f_1, \ldots, f_m$. Furthermore, this definition assumes that the eigenvalues of S^*S are all nonnegative real numbers, which we will now show must be the case.

Fact 19.6
All eigenvalues of S^*S are nonnegative.

[37] Recall earlier we defined $a_1, \ldots, a_n$ as the singular values of S, not $|a_1|, \ldots, |a_n|$. However, we will later show that the eigenvalues of S^*S must be real and nonnegative, so $a_i = |a_i|$.

Proof. Let $T = S^*S$. Then,

$$T^* = (S^*S)^* = S^*(S^*)^* = S^*S = T,$$

so T is self-adjoint. By the Spectral Theorem, T is diagonalizable under some orthonormal basis $e_1, \ldots, e_n$. Thus, $Te_i = \lambda_i e_i$ for $i = 1, 2, \ldots, n$. Then,

$$\langle e_i, S^*Se_i \rangle = \langle Se_i, Se_i \rangle = ||Se_i||^2 \geq 0$$

and

$$\langle e_i, S^*Se_i \rangle = \langle e_i, \lambda e_i \rangle = \overline{\lambda_i}.$$

Therefore, $\lambda_i \geq 0$, as desired. □

Finally, we will answer the Guiding Question posed at the start of this section.

Theorem 19.7
Suppose $S \in \mathcal{L}(V, W)$ and $\dim V \leq \dim W$. Then, there exists some orthonormal bases $e_1, \ldots, e_n$ of V and $f_1, \ldots, f_m$ of W such that

$$S(e_i) = a_i f_i$$

for $1 \leq i \leq n$ and a_i are the singular values of S.

Singular Value Decomposition. We will give a simplified version of the proof, where we assume that S is injective. The details of the general proof is in the textbook.

Suppose $S^*Sv = 0$. Then, $\langle Sv, Sv \rangle = \langle v, S^*Sv \rangle = 0$, which implies $Sv = 0$. Since S is injective, it follows that $v = 0$. Therefore, S^*S is injective.

Furthermore, we proved earlier that S^*S is self-adjoint, so S^*S is diagonalizable under some orthonormal basis $e_1, \ldots, e_n$ of V. Thus,

$$S^*S(e_i) = \lambda_i e_i$$

where $\lambda_i > 0$, because 19.6 and $\lambda \neq 0$ because S is injective. By definition, the singular values of S are $a_i = \sqrt{\lambda_i}$.

Now, let $f_i = \frac{S(e_i)}{a_i}$ for $i = 1, 2, \ldots, n$. Furthermore,

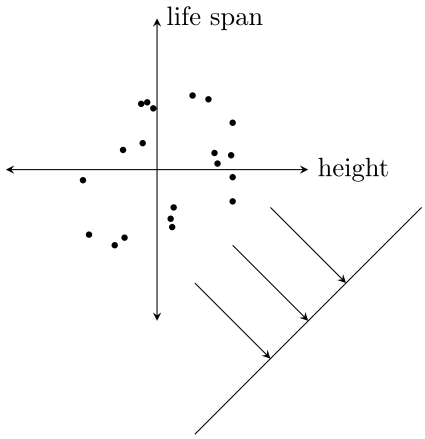
$$\begin{aligned}
\langle f_i, f_j \rangle &= \langle \tfrac{S(e_i)}{a_i}, \tfrac{S(e_j)}{a_j} \rangle \\
&= \frac{1}{a_i a_j} \langle S(e_i), S(e_j) \rangle \\
&= \frac{1}{a_i a_j} \langle e_i, S^*Se_j \rangle \\
&= \frac{\lambda_j}{a_i a_j} \langle e_i, e_j \rangle \\
&= \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases}
\end{aligned}$$

so $f_1, \ldots, f_n$ is an orthonormal list. Finally, extending $f_1, \ldots, f_n$ to an orthonormal basis gives us the desired orthonormal bases. □

19.4 Applications of Singular Values

Now, we will discuss some real-world applications of singular values.

Suppose we wish to study the relationship height and life span. We would obtain a scatter plot similar to the following, where the data is normalized such that the data is centered around 0. We wish to project this data onto a line such that the characteristics of the data is preserved as much as possible.



Then, the direction of the "best projection" is given by an eigenvector of S^*S, where $S : \mathbb{R}^2 \to \mathbb{R}^n$. The inner product space $\mathbb{R}^2$ consists of vectors representing the ordered pairs (height, life span) and the inner product space $\mathbb{R}^n$ consists of vectors representing each data point in the plot, where n is the total number of data points.

We can compute that the matrix of S^*S is a symmetric matrix such that

$$\mathcal{M}(S^*S) = \begin{pmatrix} \sum(\text{height})^2 & \sum(\text{height} \cdot \text{life span}) \\ \sum(\text{height} \cdot \text{life span}) & \sum(\text{life span})^2 \end{pmatrix}.$$

The larger the magnitude of $\sum(\text{height} \cdot \text{life span})$, the stronger the relationship between the two variables. Thus, it can be deduced that direction of the "best projection" is given by the eigenvector ***corresponding to the largest eigenvalue*** of S^*S,

20 Polar Decomposition, Generalized Eigenspaces, and Nilpotent Operators

20.1 Review

Recall the properties of the various classes of operators that we have discussed over the past few lectures:

- Self-adjoint operators: $T = T^*$ (see Definition 17.8),
- Normal operators: $T^*T = TT^*$ (see Definition 18.8),
- Isometries: $T^*T = I$ (see Definition 19.1 and Theorem 19.4).

20.2 Positive Operators

Before we continue, we must first introduce a new class of operators.

Definition 20.1 (Positive operators)
An operator $T \in \mathcal{L}(V)$ is **positive** if T is self-adjoint and all eigenvalues of T are nonnegative.

We know that all self-adjoint operators have only real eigenvalues; thus, positive operators is a subset of self-adjoint operators.

20.3 Relationship Between Operators and Complex Numbers

If $T \in \mathcal{L}(V)$ is self-adjoint, then T and T^* must commute because $T = T^*$, so T is also normal. Similarly, if T is an isometry, then $T^*T = TT^* = I$, so T is normal. Therefore, self-adjoint operators and isometries are both subsets of normal operators. In other words, we have the relation

$$\text{Isomtries} \subset \text{Normal} \supset \text{Self-adjoint} \supset \text{Positive}.$$

Now, consider the case where $\dim V = 1$, so each operator T can be represented by a 1×1 matrix, which is essentially a number. In this scenario, a complex number z corresponds to an operator T, and $\overline{z}$ corresponds to T^*. Then, we can determine what each class of operators in $\mathcal{L}(V)$ corresponds to in $\mathbb{C}$:

- Normal operators: $z\overline{z} = \overline{z}z$ holds for all z, so normal operators correspond to $\mathbb{C}$,
- Self-adjoint operators: $z = \overline{z}$ holds for real z, so self-adjoint operators correspond to $\mathbb{R}$,
- Positive operators: self-adjoint operators with nonnegative eigenvalues correspond to nonnegative real numbers, or $\mathbb{R}_{\geq 0}$,
- Isometries: for any $v \in V$, it must be that $||v|| = ||z \cdot v||$, which holds for all $|z| = 1$, so isometries correspond to the unit circle in $\mathbb{C}$.

Now, it follows that that our above relation about operators is parallel to the following relation about complex numbers:

$$\text{Unit circle} \subset \mathbb{C} \supset \mathbb{R} \supset \mathbb{R}_{\geq 0}.$$

This is more than an analogy between $\mathcal{L}(V)$ and $\mathbb{C}$, but rather the special case of opeartors where $\dim V = 1$.

Exercise 20.2
The unit circle and $\mathbb{R}$ only intersect at two points (1 and -1). Can you use this fact to find all operators which are both isometries and self-adjoint?

20.4 Polar Decomposition

Note that every nonzero complex number z can be written uniquely as

$$z = z_1 r,$$

where $|z_1| = 1$ and $r > 0$ (in particular, $z_1 = e^{i\theta}$ and $r = |z|$).

In terms of operators, z_1 corresponds to an isometry and r corresponds to a positive operator. Thus, for any invertible[38] operator $T \in \mathcal{L}(V)$, we can guess that T can be written as the product of an isometry and a positive operator. This leads us to the following result.

Theorem 20.3 (Polar decomposition)
Suppose V is an inner product space over $\mathbb{C}$ and $T \in \mathcal{L}(V)$ is invertible. Then, there is a unique way of expressing

$$T = SP,$$

where $S \in \mathcal{L}(V)$ is an isometry and $P \in \mathcal{L}(V)$ is positive.

Proof. We will find a formula for P. Since $T = SP$, taking the adjoint of both sides gives $T^* = P^*S^*$. Multiplying these two equations gives

$$T^*T = P^*S^*SP = P^*P = P^2,$$

where the second equality holds because S is an isometry and the last equality holds because P is self-adjoint.

Since T^*T is self-adjoint, the Spectral Theorem tells us that there is an orthonormal basis $e_1, \ldots, e_n$ of V such that $T^*T(e_i) = \lambda_i e_i$ for $\lambda_i \in \mathbb{R}$. In fact,

$$\lambda_i = \overline{\lambda_i} = \langle e_i, \lambda_i e_i \rangle = \langle e_i, T^*Te_i \rangle = \langle Te_i, Te_i \rangle \geq 0,$$

so all λ_i are nonnegative. Thus, T^*T is a positive operator. Now, let $P(e_i) = \sqrt{\lambda_i} e_i$. It is clear that $P^2 = T^*T$ and all $\sqrt{\lambda_i} \geq 0$, so P is positive.

Since T is invertible, T^* is also invertible, so T^*T is invertible. It follows that all λ_i are strictly positive, so all $\sqrt{\lambda_i}$ are also stricly positive, so P is invertible. Now, let $S = TP^{-1}$. It remains to show that S is an isometry. We know that

$$S^* = (TP^{-1})* = (P^{-1})^*T^* = P^{-1}T^*,$$

where the last equality holds because P is self-adjoint, so P^* is also self-adjoint. Then,

$$S^*S = P^{-1}T^*TP^{-1} = P^{-1}P^2P^{-1} = I,$$

so S is an isometry, as desired.

To prove uniqueness, we must show that $P^2 = T^*T$ uniquely defines P. In fact, it is proved in the textbook that every positive operator has a unique square root, so P is unique. □

The following result is a more general form of polar decomposition, which we will not prove.

Theorem 20.4 (General polar decomposition)
Suppose $T \in \mathcal{L}(V, W)$ and $\dim V \leq \dim W$. Then, there exists an isometric embedding $S \in \mathcal{L}(V, W)$ such that

$$T = S\sqrt{T^*T}.$$

Note that $T^*T \in \mathcal{L}(V)$ is a positive operator, so $\sqrt{T^*T}$ is uniquely determined.

20.5 Generalized Eigenspaces

Now, we will shift our discussion from inner product spaces back to the general setting of vector spaces. For the remainder of this chapter, we will focus our attention on complex vector spaces.

Recall that the eigenspaces of $T \in \mathcal{L}(V)$ are defined by $E(\lambda, T) = \text{Null}(T - \lambda I)$, and T is diagonalizable if and only if $V = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$. However, we have seen that not all operators are diagonlizable.

[38] Note that z being nonzero in $\mathbb{C}$ corresponds to T being invertible in $\mathcal{L}(V)$.

Guiding Question

Suppose $T \in \mathcal{L}(V)$. Even if T is not diagonalizable, can we still have a decomposition of V into a direct sum of subspaces such that

$$V = (\quad)_1 \oplus \cdots \oplus (\quad)_m,$$

where $(\quad)_i$ is a subspace related to λ_i?

Our goal is to show that the answer to the above question is yes.

Example 20.5

Suppose $T \in \mathcal{L}(V)$ such that

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

First, we will find the eigenvalues of T. It helps to see that this matrix is block upper-triangular:

$$\mathcal{M}(T) = \left(\begin{array}{cc|c} 0 & 0 & 1 \\ 1 & 0 & 0 \\ \hline 0 & 0 & 1 \end{array}\right).$$

It follows that $U = \operatorname{span}(e_1, e_2)$ is T-invariant and $T/U \in \mathcal{L}(V/U)$ has matrix 1 under $\overline{e_3}$, where the notation $\overline{e_3}$ represents the projection of e_3 onto V/U. Now, the eigenvalues of T are the union of the eigenvalues of $T|_U$ and T/U.

First, $T|_U$ has matrix $\mathcal{M}(T|_U) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, which has eigenvalues that satisfy the equation

$$\det \begin{pmatrix} \lambda & 0 \\ 1 & \lambda \end{pmatrix} = \lambda^2 = 0,$$

so the set of eigenvalues of $T|_U$ is $\{0\}$. From $\mathcal{M}(T|_U)$, we can deduce that

$$Te_1 = e_2 \quad \text{and} \quad Te_2 = 0.$$

Thus, e_2 is an eigenvector of T, while e_1 is not. However,

$$T^2 e_1 = T(Te_1) = Te_2 = 0.$$

Thus, while e_1 is not an eigenvector of T, it is an eigenvector of T^2. This is a special case of a ***generalized eigenspace***, where

$$G(0, T) = \operatorname{span}(e_1, e_2)$$

such that every $v \in G(0, T)$ satisfies $T^2 v = 0$. This can also be seen by the fact that

$$\mathcal{M}((T|_U)^2) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

so $(T|_U)^2 = 0$.

Now, since T/U has 1×1 matrix 1, it follows that T/U is the identity operator on V/U. Thus, the eigenvalues of T/U are $\{1\}$. Therefore, the eigenvalues of T are $\{0, 1\}$.

If we compute the eigenvectors of T corresponding to eigenvalue 1, we find that there is a unique eigenvector v_1 (up to scalar multiplication). Thus, $E(1, T) = \operatorname{span}(v_1)$.

It follows that we can decompose V into

$$V = G(0, T) \oplus E(1, T)$$

because e_1, e_2, v_1 are linearly independent and $\dim V = 3 = \dim G(0, T) + \dim E(1, T)$.

We will soon see that $G(1,T) = E(1,T)$ in the above example. Thus, we can guess that we can always decompose V into a direct sum of generalized eigenspaces of T.

First, we must give a formal definition of generalized eigenspaces.

Definition 20.6 (generalized eigenspace)
Suppose $T \in \mathcal{L}(V)$. The **generalized eigenspace** of T corresponding to λ is defined by

$$G(\lambda, T) = \{v \in V | (T - \lambda I)^i v = 0 \text{ for some } i > 0\}.$$

It is easy to verify that $G(\lambda, T)$ is a subspace. Furthermore, recall that the eigenspace of T corresponding to λ is defined by

$$E(\lambda, T) = \{v \in V | (T - \lambda I)v = 0\}.$$

Thus, it is clear that $E(\lambda, T) \subset G(\lambda, T)$.

In the definition above, note that there are no constraints on the value of i such that $(T - \lambda I)^i v = 0$. The next result will give an upper bound on i.

Lemma 20.7
Suppose $T \in \mathcal{L}(V)$ and $v \in G(0,T)$. Let $n = \dim V$. Then, $T^n v = 0$.

Proof. Suppose $v \in \text{Null}(T^k)$. Then, $T^{k+1}v = T(T^k v) = T(0) = 0$, so $v \in \text{Null}(T^{k+1})$. Thus, $\text{Null}(T^k) \subset \text{Null}(T^{k+1})$, which implies that

$$\{0\} \subset \text{Null}(T) \subset \text{Null}(T^2) \subset \cdots \subset \text{Null}(T^k) \subset \text{Null}(T^{k+1}) \subset \cdots \subset V.$$

Now, note that $\text{Null}(T), \text{Null}(T^2), \ldots, \text{Null}(T^i), \ldots$ cannot keep growing infinitely, because each null space is a subspace of the finite-dimensional vector space V. Let i be the smallest integer such that $\text{Null}(T^i) = \text{Null}(T^{i+1})$. Suppose $v \in \text{Null}(T^{i+2})$. Then,

$$T^{i+2}v = T^{i+1}(Tv) = 0,$$

so $Tv \in \text{Null}(T^{i+1})$. Since $\text{Null}(T^i) = \text{Null}(T^{i+1})$, it follows that $Tv \in \text{Null}(T^i)$, so

$$T^i(Tv) = T^{i+1}v = 0.$$

Therefore, $v \in \text{Null}(T^{i+1})$, so $\text{Null}(T^{i+2}) \subset \text{Null}(T^{i+1})$. Since we already know that $\text{Null}(T^{i+i}) \subset \text{Null}(T^{i+2})$, it follows that $\text{Null}(T^{i+i}) = \text{Null}(T^{i+2})$. Using the same logic for higher powers of T, it follows that

$$\text{Null}(T^i) = \text{Null}(T^{i+1}) = \text{Null}(T^{i+2}) = \text{Null}(T^{i+3}) = \cdots.$$

Thus, we have the relation

$$\{0\} \subsetneq \text{Null}(T) \subsetneq \text{Null}(T^2) \subsetneq \cdots \subsetneq \text{Null}(T^i) = \text{Null}(T^{i+1}) = \text{Null}(T^{i+2}) = \cdots \subset V.$$

This implies that

$$0 < \dim \text{Null}(T) < \dim \text{Null}(T^2) < \cdots < \dim \text{Null}(T^i),$$

so $\dim \text{Null}(T^i) \geq i$. Furthermore, $\text{Null}(T^i) \subset V$, so $\dim \text{Null}(T^i) \leq n$. It follows that $n \geq i$, so

$$\text{Null}(T^i) = \cdots = \text{Null}(T^n) = \text{Null}(T^{n+1}) = \cdots.$$

Now, let $v \in G(0,T)$. By definition, there exists some $j > 0$ such that $T^j v = 0$, so $v \in \text{Null}(T^j)$. If $j < n$, then $\text{Null}(T^j) \subset \text{Null}(T^n)$. If $j \geq n$, then $\text{Null}(T^j) = \text{Null}(T^n)$. In either scenario, it follows that $v \in \text{Null}(T^n)$, so $T^n v = 0$, as desired. □

We can generalize the above result to all generalized eigenspaces.

Theorem 20.8
Suppose $T \in \mathcal{L}(V)$ and $v \in G(\lambda, T)$. Let $n = \dim V$. Then, $(T - \lambda I)^n v = 0$.

Proof. The proof is the same as Lemma 20.7 by replacing T with $T - \lambda I$. □

20.6 Nilpotent Operators

Now, we will introduce nilpotent operators.

Definition 20.9 (nilpotent)
An operator $T \in \mathcal{L}(V)$ is called **nilpotent** if $G(0,T) = V$.

In other words, T is a nilpotent operator if for any $v \in V$, there exists some $i > 0$ such that $T^i v = 0$.

Example 20.10
Suppose the matrix of $T \in \mathcal{L}(V)$ is the strictly upper-triangular matrix

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

We will show that T is nilpotent.

The matrix of T implies that

$$\begin{aligned} Te_1 &= 0 \\ Te_2 &\in \operatorname{span}(e_1) \\ Te_3 &\in \operatorname{span}(e_1, e_2) \\ &\vdots \\ Te_n &\in \operatorname{span}(e_1, \dots, e_{n-1}). \end{aligned}$$

This implies that $Tv = T(c_1e_1 + \cdots + c_ne_n) \in \operatorname{span}(e_1, \dots, e_{n-1})$, so $\operatorname{Range}(T) \subset \operatorname{span}(e_1, \dots, e_{n-1})$. By similar logic, we see that $\operatorname{Range}(T^2) \subset \operatorname{span}(e_1, \dots, e_{n-2})$. Continuing this pattern, we see that $\operatorname{Range}(T^{n-1}) \subset \operatorname{span}(e_1)$, so $\operatorname{Range}(T^n) = \{0\}$. Therefore, $T^n = 0$, so T is nilpotent.

The next result will further illuminate the properties of nilpotent operators.

Proposition 20.11
Suppose $T \in \mathcal{L}(V)$. The following are equivalent:

1. T is nilpotent,
2. All eigenvalues of T are 0,
3. $\mathcal{M}(T)$ is strictly upper-triangular under some basis of V.

Proof. First, we will show that (1) implies (2). Suppose λ is an eigenvalue of T. Then, there exists some v such that $Tv = \lambda v$. Since T is nilpotent, it follows that $T^i v = 0$ for some $i > 0$. Then,

$$T^i v = \lambda^i v = 0,$$

so $\lambda = 0$. Thus, all eigenvalues of T are 0.

Now, we will show that (2) implies (3). By Corollary 13.5, there exists a basis $v_1, \dots, v_n$ of V under which $\mathcal{M}(T)$ is upper-triangular. Furthermore, by Proposition 12.11, the diagonal entries of $\mathcal{M}(T)$ are exactly the eigenvalues of T. Since all the eigenvalues are 0, it follows that $\mathcal{M}(T)$ is strictly upper-triangular.

Finally, we have already shown that (3) implies (1) in Example 20.10. Thus, all three statements are equivalent. □

Example 20.10 and Proposition 20.11 imply that if $T \in \mathcal{L}(V)$ is nilpotent, then $T^n = 0$. This can also be seen by Lemma 20.7 in combination with the fact that $V = G(0,T)$.

In a later lecture, we will study the Jordan form of operators, which will tell us that for any nilpotent operator $T \in \mathcal{L}(V)$, there exists a basis under which $\mathcal{M}(T)$ has 0's everywhere except possibly the line directly above the diagonal (called the ***superdiagonal***), which consists of 0's and 1's:

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

21 Generalized Eigenspaces (continued)

21.1 Review

Last time, we introduced the notion of generalized eigenspaces:

$$G(\lambda, T) = \{v \in V | (T - \lambda I)^i v = 0 \text{ for some } i > 0\}.$$

Furthermore, we showed in Theorem 20.8 that i is upper-bounded by $\dim V$, so $G(\lambda, T) = \text{Null}((T - \lambda I)^{\dim V})$.

21.2 Multiplicity of an Eigenvalue

Now, we will introduce multiplicity.

Definition 21.1 (multiplicty)
Suppose $T \in \mathcal{L}(V)$. The **multiplicity** of an eigenvalue λ of T is $\dim G(\lambda, T)$.

Since the generalized eigenspace $G(\lambda, T)$ always contains the eigenspace $E(\lambda, T)$, it follows that the multiplicity of λ of T is nonzero if λ is an eigenvalue. The following result will show that if λ is not an eigenvalue, then the multiplicity of λ is 0.

Fact 21.2
Suppose $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{C}$. If λ is not an eigenvalue of T, then $G(\lambda, T) = \{0\}$.

Proof. Suppose there exists $v \in G(\lambda, T)$ such that $v \neq 0$. Then, let i be the minimal exponent such that $(T - \lambda I)^i v = 0$. Let $u = (T - \lambda I)^{i-1} \neq 0$. It follows that

$$(T - \lambda I)u = (T - \lambda I)^i v = 0,$$

so u is an eigenvector with corresponding eigenvalue λ. However, this is a contradiction because we assumed that λ is not an eigenvalue of T. Therefore, $G(\lambda, T) = \{0\}$. □

This explains why we do not define the notion of a generalized eigenvalue, since that would be identical to our previous definition of eigenvalues.

21.3 Decomposition of an Operator

We will now prove the main result of generalized eigenspaces.

Theorem 21.3
Suppose $T \in \mathcal{L}(V)$ and $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues of T. Then,

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T).$$

Proof. First, we will show that $G(\lambda_1, T) + \cdots + G(\lambda_m, T)$ is a direct sum. It is sufficient to show that if $v_i \in G(\lambda_i, T)$ and $v_1 + \cdots + v_m = 0$, then each $v_i = 0$. Suppose there exists some $v_i \neq 0$. Without loss of generality, assume $v_1 \neq 0$. Let $n = \dim V$. It is clear that

$$(T - \lambda_2 I)^n v_2 = (T - \lambda_3 I)^n v_3 = \cdots = (T - \lambda_m I)^n v_m = 0.$$

Now, let $S = (T - \lambda_2 I)^n (T - \lambda_3 I)^n \cdots (T - \lambda_m I)^n$. It follows that $Sv_2 = Sv_3 = \cdots = Sv_m = 0$.

Now, we will show that $Sv_1 \neq 0$. Let i be the minimal exponent such that $(T - \lambda I)^i v_1 = 0$. Then, let $u = (T_\lambda I)^{i-1} \neq 0$. It follows that $Tu = \lambda_1 u$, so

$$Su = (T - \lambda_2 I)^n \cdots (T - \lambda_m I)^n = (\lambda_1 - \lambda_2)^n \cdots (\lambda_1 - \lambda_m)^n u \neq 0,$$

since $\lambda_1, \ldots, \lambda_m$ are all distinct. Noe, let $P = S(T - \lambda_1 I)^{i-1}$. Then,

$$\begin{aligned} P(v_1 + \cdots + v_m) &= S(T - \lambda_1 I)^{i-1} v_1 + \cdots S(T - \lambda_1 I)^{i-1} v_m \\ &= Su + (T - \lambda_1 I)^{i-1} S v_2 + \cdots + (T - \lambda_1 I)^{i-1} S v_m \\ &= Su \\ &\neq 0. \end{aligned}$$

However, $P(v_1 + \cdots + v_m) = P(0) = 0$, which is a contradiction. Therefore, $G(\lambda_1, T) + \cdots + G(\lambda_m, T)$ is a direct sum.

Now, we will show that $G(\lambda_1, T), \ldots, G(\lambda_m, T)$ span V. We will do this using induction on $\dim V$. Let $n = \dim V$. It is clear that the desired result holds for $n = 1$. Now, assume $n > 1$ and the desired result holds for all vector spaces of smaller dimension. Let $U = \text{Range}((T - \lambda_1 I)^n)$. We will show that

$$V = G(\lambda_1, T) \oplus U.$$

Since $\dim V = \dim \text{Null}((T - \lambda_1 I)^n) + \dim \text{Range}((T - \lambda_1 I)^n) = \dim G(\lambda_1, T) + \dim U$, it is sufficient to show that $G(\lambda_1, T) + U$ is a direct sum. Suppose $v \in G(\lambda_1, T) \cup U$. It follows that $(T - \lambda_1 I)^n v = 0$ (because $v \in G(\lambda_1, T)$) and $v = (T - \lambda_i I)^n u$ for some $u \in V$ (because $v \in U$). This implies that $(T - \lambda_1 I)^{2n} u = 0$, so $u \in G(\lambda_1, T)$. By Theorem 20.8, it follows that $(T - \lambda_i I)^n u = 0$. Since we defined $v = (T - \lambda_i I)^n u$, we know that $v = 0$, so $G(\lambda_1, T) + U$ is a direct sum.

Now, suppose $v \in U$. Then, there exists some $u \in V$ such that $v = (T - \lambda_1 I)^n u$. It follows that

$$Tv = T((T - \lambda_1 I)^n u) = (T - \lambda_1 I)^n (Tu),$$

so $Tv \in \text{Range}((T - \lambda_1 I)^n) = U$. Thus, U is T-invariant. Because $\dim U < \dim V$, we can apply the induction hypothesis to $T|_U \in \mathcal{L}(U)$, which says that $U = G(\lambda_1', T|_U) \oplus \cdots \oplus G(\lambda_{m'}', T|_U)$, where $\lambda_1', \ldots, \lambda_{m'}'$ are the eigenvalues of $T|_U$. However, eigenvalues of $T|_U$ are also eigenvalues of T, so

$$\begin{aligned} U &= G(\lambda_1', T|_U) + \cdots + G(\lambda_{m'}', T|_U) \\ &\subset G(\lambda_1', T) + \cdots + G(\lambda_{m'}', T) \\ &\subset G(\lambda_1, T) + \cdots + G(\lambda_m, T). \end{aligned}$$

It follows that

$$\begin{aligned} V &= G(\lambda_1, T) + U \\ &= G(\lambda_1, T) + (G(\lambda_1, T) + \cdots + G(\lambda_m, T)) \\ &= G(\lambda_1, T) + \cdots + G(\lambda_m, T). \end{aligned}$$

It follows that $G(\lambda_1, T), \ldots, G(\lambda_m, T)$ span V, and we are done. □

21.4 Generalized Eigenspaces for Restriction and Quotient Operators

Suppose $T \in \mathcal{L}(V)$ and U is a T-invariant subspace of V. Then, we have the restriction operators $T|_U \in \mathcal{L}(U)$ and the quotient operator $T/U \in \mathcal{L}(V/U)$.

First, we will consider the relationship between $G(\lambda, T|_U)$ and $G(\lambda, T)$. Suppose $v \in G(\lambda, T|_U)$. It follows that $(T_|U - \lambda I)^i v = 0$ for some $i > 0$. However, $(T_|U - \lambda I)^i v = (T - \lambda I)^i v = 0$, so $v \in G(\lambda, T)$. Therefore, $G(\lambda, T|_U) \subset G(\lambda, T)$.

The relationship between $G(\lambda, T/U)$ and $G(\lambda, T)$ is less obvious. Recall the quotient map π (see Definition 12.7). We will show that π maps $G(\lambda, T)$ to $G(\lambda, T/U)$. Suppose $v \in G(\lambda, T)$, so $(T - \lambda I)^i v = 0$ for some $i > 0$. Then,

$$\pi((T - \lambda I)^i v) = (T/U - \lambda I)^i (\pi(v)) = 0,$$

so $\pi(v) \in G(\lambda, T/U)$.

Now, let $\pi_\lambda : G(\lambda, T) \to G(\lambda, T/U)$ such that $\pi_\lambda(v) = \pi(v)$ for any $v \in G(\lambda, T)$. Recall that $\text{Null}(\pi) = U$. Then,

$$\text{Null}(\pi_\lambda) = \{v \in G(\lambda, T),\ \pi(v) = 0\} = \{v \in G(\lambda, T) \cup U\} = G(\lambda, T|_U),$$

which shows how $G(\lambda, T/U)$ and $G(\lambda, T|_U)$ are related.

Now, we will show that the map $\pi_\lambda : G(\lambda, T) \to G(\lambda, T/U)$ is surjective.

Lemma 21.4
The map π_λ (as defined above) is surjective.

Proof. We know that

$$\dim G(\lambda, T/U) \geq \dim \operatorname{Range}(\pi_\lambda) = \dim G(\lambda, T) - \dim G(\lambda, T|_U),$$

where the inequality follows from the fact that $\operatorname{Range}(\pi_\lambda) \subset G(\lambda, T/U)$ and the equality follows from the rank-nullity theorem and $\operatorname{Null}(\pi_\lambda) = G(\lambda, T|_U)$. Summing this inequality over all λ, we get

$$\dim V/U \geq \dim V - \dim U.$$

However, we know that $\dim V/U = \dim V - \dim U$, which forces $\dim G(\lambda, T/U) = \dim G(\lambda, T) - \dim G(\lambda, T|_U)$ for all λ. Therefore, $\dim G(\lambda, T/U) = \operatorname{Range}(\pi_\lambda)$, so π_λ is surjective. □

Now, we are ready to prove the main result of this section.

Theorem 21.5
Suppose $T \in \mathcal{L}(V)$ and U is a T-invariant subspace of V. Then,

$$\operatorname{mult}_\lambda(T) = \operatorname{mult}_\lambda(T|_U) + \operatorname{mult}_\lambda(T/U),$$

where $\operatorname{mult}_\lambda(T)$ represents the multiplicity of λ of T.

Proof. Since π_λ is surjective, it follows that $\operatorname{Range}(\pi_\lambda) = G(\lambda, T/U)$. It follows that

$$\dim G(\lambda, T) = \dim \operatorname{Null}(\pi_l l) + \dim \operatorname{Range}(\pi_\lambda) = \dim G(\lambda, T|_U) + \dim G(\lambda, T/U).$$

In other words, $\operatorname{mult}_\lambda(T) = \operatorname{mult}_\lambda(T|_U) + \operatorname{mult}_\lambda(T/U)$, as desired. □

However, note that this does NOT necessarily the case that $\dim E(\lambda, T) = \dim E(\lambda, T|_U) + \dim E(\lambda, T/U)$.

Example 21.6
Consider $T \in \mathcal{L}(V)$ such that

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

Let $U = \operatorname{span}(e_1)$. It is clear that $E(0, T) = E(0, T|_U) = \operatorname{span}(e_1) = U$. Also, it follows that $\mathcal{M}(T/U)$ is the lower-right block of $\mathcal{M}(T)$, which is the 1×1 zero matrix. Thus, $\dim E(0, T/U) = 1$. Therefore,

$$\dim E(0, T) = 1 \neq 1 + 1 = \dim E(0, T|_U) + \dim E(0, T/U).$$

In general, this is because Lemma 21.4 does not hold for non-generalized eigenspaces. For instance, the map $\pi_\lambda : E(0, T) \to E(0, T/U)$ would map all $v \in E(0, T)$ to the zero vector because $E(0, T) = U$, so π_λ is not surjective.

21.5 Eigenvalues of Upper-Triangular Matrices

Suppose $T \in \mathcal{L}(V)$ and the matrix representation of T is the upper-triangular matrix

$$\mathcal{M}(T) = \begin{pmatrix} a_1 & & * \\ & \ddots & \\ 0 & & a_n \end{pmatrix}.$$

By Proposition 12.11, we know that the eigenvalues of T are the values along the diagonal. The next result will tell us how to determine the multiplicities of each eigenvalue based on the matrix.

Proposition 21.7
Suppose $T \in \mathcal{L}(V)$ and the matrix of T is upper-triangular. The multiplicity of $\lambda \in \mathbb{C}$ of T is equivalent to the number of times λ appears on the diagonal of $\mathcal{M}(T)$.

Proof. We will prove the desired result by induction on $\dim V$. Let $n = \dim V$ and $\mathcal{M}(T)$ have diagonal entries $a_1, \ldots, a_n$. The result is obvious for $n = 1$. For $n > 1$, we can partition $\mathcal{M}(T)$ into

$$\left(\begin{array}{c|ccc} a_1 & & * & \\ \hline & a_2 & & * \\ 0 & & \ddots & \\ & 0 & & a_n \end{array}\right).$$

Then, $U = \operatorname{span}(e_1)$ is T-invariant, so $\mathcal{M}(T|_U)$ is the upper-left block of $\mathcal{M}(T)$ and $\mathcal{M}(T/U)$ is the lower-right block. By the applying the induction hypothesis on $T|_U$ and T/U, we know that

$$\begin{aligned} \operatorname{mult}_\lambda(T|_U) &= \# \text{ of times } \lambda \text{ appears in } (a_1) \\ \operatorname{mult}_\lambda(T/U) &= \# \text{ of times } \lambda \text{ appears in } (a_2, \ldots, a_n). \end{aligned}$$

Finally, by Theorem 21.5,

$$\begin{aligned} \operatorname{mult}_\lambda(T) &= \operatorname{mult}_\lambda(T|_U) + \operatorname{mult}_\lambda(T/U) \\ &= (\# \text{ of times } \lambda \text{ appears in } (a_1)) + (\# \text{ of times } \lambda \text{ appears in } (a_2, \ldots, a_n)) \\ &= \# \text{ of times } \lambda \text{ appears in } (a_1, \ldots, a_n), \end{aligned}$$

as desired. □

Example 21.8
Suppose $T \in \mathcal{L}(V)$ and the matrix of T is the upper-triangular matrix

$$\mathcal{M}(T) = \begin{pmatrix} 0 & 3 & 4 \\ & 0 & 5 \\ & & 1 \end{pmatrix}.$$

In this example, it is not hard to see that the upper-left 2×2 block is a nilpotent matrix, so $G(0, T) = \operatorname{span}(e_1, e_2)$. Thus, the multiplicity of $\lambda = 0$ is 2.

Now, suppose the matrix of T is

$$\mathcal{M}(T) = \begin{pmatrix} 0 & 3 & 4 \\ & 1 & 5 \\ & & 0 \end{pmatrix}.$$

In this example, it is not obvious what $G(0, T)$ is (for instance, you should verify that $G(0, T) \neq \operatorname{span}(e_1, e_3)$). Nevertheless, Proposition 21.7 tells us that that the multiplicity of $\lambda = 0$ is 2.

However, what if $\mathcal{M}(T)$ is not an upper-triangular matrix, but rather a block upper-triangular matrix?

Guiding Question
Suppose

$$\mathcal{M}(T) = \left(\begin{array}{c|c} A & * \\ \hline 0 & A \end{array}\right),$$

where A is an arbitrary square matrix. Then, how are $\operatorname{mult}_\lambda(T)$ and $\operatorname{mult}_\lambda(A)$ related?

Answer. *Let A be an $n \times n$ matrix. Then, it follows that $U = \operatorname{span}(e_1, \ldots, e_n)$ is T-invariant and $\mathcal{M}(T|_U) = \mathcal{M}(T/U) = A$. We will show that this implies that $\operatorname{mult}_\lambda(T|_U) = \operatorname{mult}_\lambda(T/U)$.*

Let $f_1, \ldots, f_n$ be a basis of U such that $\mathcal{M}(T|_U) = A$. Similarly, let $g_1, \ldots, g_n$ be a basis of V/U such that $\mathcal{M}(T/U) = A$. Define the linear map $S : U \to V/U$ such that $S(e_i) = f_i$. Since S maps a basis to a basis, it *follows that S is an isomorphism. Then, it follows that*

$$S(T_1 v) = T_2(Sv).$$

The reader should verify that S sends $G(\lambda, T|_U)$ to $G(\lambda, T/U)$ and the two generalized eigenspaces are isomorphic. Thus, $G(\lambda, T|_U)$ to $G(\lambda, T/U)$ have the same dimension, so $\operatorname{mult}_\lambda(T|_U) = \operatorname{mult}_\lambda(T/U)$.

Finally,

$$\begin{aligned}
\operatorname{mult}_\lambda(T) &= \operatorname{mult}_\lambda(T|_U) + \operatorname{mult}_\lambda(T/U) \\
&= \operatorname{mult}_\lambda(A) + \operatorname{mult}_\lambda(A) \\
&= 2 \operatorname{mult}_\lambda(A).
\end{aligned}$$

The next result gives a useful matrix form of any operator.

Theorem 21.9

Suppose $T \in \mathcal{L}(V)$. Let T have distinct eigenvalues $\lambda_1, \ldots, \lambda_m$ with multiplicities $d_1, \ldots, d_m$. Then, there exists a basis of T such that

$$\mathcal{M}(T) = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix},$$

where each A_i is a $d_i \times d_i$ upper-triangular matrix of the form

$$A_i = \begin{pmatrix} \lambda_i & & * \\ & \ddots & \\ 0 & & \lambda_i \end{pmatrix}.$$

Proof. For each $T|_{G(\lambda_i, T)}$, there exists a basis of $G(\lambda_i, T)$ such that $\mathcal{M}(T|_{G(\lambda_i, T)})$ is upper-triangular. By Proposition 21.7, $\mathcal{M}(T|_{G(\lambda_i, T)})$ must be of the form

$$\begin{pmatrix} \lambda_i & & * \\ & \ddots & \\ 0 & & \lambda_i \end{pmatrix},$$

since λ_i is the only eigenvalue of $T|_{G(\lambda_i, T)}$. Now, since $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$, concatenating all these bases of $G(\lambda_i, T)$ gives a basis of V such that $\mathcal{M}(T)$ is of the desired form. □

We will expand on the fact that $T \in \mathcal{L}(V)$ can be expressed as a block diagonal matrix when we introduce Jordan form in a later lecture.

22 Characteristic Polynomial and Jordan Form

22.1 Review

Last time, we discussed how any complex vector space V can be decomposed into the generalized eigenspaces of any operator $T \in \mathcal{L}(V)$ (see Theorem 21.3). Also, we defined the multiplicity of λ of T as the dimension of $G(\lambda, T)$.

22.2 Characteristic Polynomial

Now, we will introduce the characteristic polynomial.

Definition 22.1
Suppose $T \in \mathcal{L}(V)$. Let $\lambda_1, \dots, \lambda_m$ denote the distinct eigenvalues of T with multiplicities $n_1, \dots, n_m$. Then, the **characteristic polynomial** of T is

$$(x - \lambda_1)^{n_1} \cdots (x - \lambda_m)^{n_m}.$$

Note that given the characteristic polynomial of any operator $T \in \mathcal{L}(V)$, we are able to determine all the eigenvalues of T with their multiplicities by finding the roots of the polynomial.

Example 22.2
Suppose $T \in \mathcal{L}(V)$ and the characteristic polynomial of T is $x^3 - 2$. Then, the eigenvalues of T are the solutions to the equation $x^3 - 2 = 0$, which happen to be

$$\sqrt[3]{2},\ \sqrt[3]{2}\omega,\ \sqrt[3]{2}\overline{\omega},$$

where $\omega = \frac{-1+\sqrt{3}i}{2}$. In this example, all three eigenvalues have multiplicity 1.

The following result gives an important property of characteristic polynomials.

Theorem 22.3 (Cayley-Hamilton Theorem)
Suppose $T \in \mathcal{L}(V)$. Let $p(x)$ denote the characteristic polynomial of T. Then, $p(T)$ is the zero operator on V.

Proof. Let $\lambda_1, \dots, \lambda_m$ denote the distinct eigenvalues of T with multiplicities $n_1, \dots, n_m$. Denote $N_i = (T - \lambda_i I)|_{G(\lambda_i, T)} \in \mathcal{L}(G(\lambda_i, T))$. For all $v \in G(\lambda_i, T)$, we know by Theorem 20.8 that $N_i^{n_i} v = 0$, so N_i is nilpotent and $N_i^{n_i} = 0$. Therefore,

$$(T - \lambda_i I)^{n_i} v_i = 0$$

for all $v_i \in G(\lambda_i, T)$. Since $(T - \lambda_i I)^{n_i}$ is a factor of $p(T)$, it follows that $p(T) v_i = 0$. Thus, $G(\lambda_i, T) \subset \text{Null}(p(T))$. Since this holds for arbitrary λ_i, it follows that $G(\lambda_1, T), \dots, G(\lambda_m, T)$ are all contained in $\text{Null}(p(T))$. However, since $G(\lambda_1, T), \dots, G(\lambda_m, T)$ span V, it follows that $\text{Null}(p(T)) = V$. Therefore, $p(T) = 0$. $\square$

Example 22.4
Suppose $T \in \mathcal{L}(V)$ and the matrix of T is

$$\mathcal{M}(T) = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

We claim (without proof) that the characteristic polynomial of T is $x^2 - (a+d)x + (ad - bc)$.[a] Then, calculating the matrix of $p(T)$ gives

$$\begin{aligned} \mathcal{M}(p(T)) &= \mathcal{M}(T^2 - (a+d)T + (ad-bc)I) \\ &= \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} - \begin{pmatrix} (a+d)a & (a+d)b \\ (a+d)c & (a+d)d \end{pmatrix} + \begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \end{aligned}$$

the last equality you should verify yourself. Thus, $p(T) = 0$, as suggested by the Cayley-Hamilton Theorem.

[a] We will see why this is the case after we discuss trace and determinant in a later lecture.

22.3 Minimal Polynomial

In this section, we will introduce a polynomial that is closely related to the characteristic polynomial. Before that, we will need the following definition.

Definition 22.5 (monic polynomial)
A **monic polynomial** is a polynomial whose leading coefficient equals 1.

Example 22.6
Consider the following polynomials:

- $T^3 + T + 2$ is a monic polynomial,
- $T - 1$ is a monic polynomial,
- $2T + 1$ is NOT a monic polynomial.

Now, we can define the minimal polynomial.

Definition 22.7 (minimal polynomial)
Suppose $T \in \mathcal{L}(V)$. The **minimal polynomial** of T is the monic polynomial $m(x)$ of lowest possible degree such that $m(T) = 0$.

While the characteristic polynomial $p(x)$ and the minimal polynomial $m(x)$ have the same condition of $p(T) = m(T) = 0$, they are not necessarily equivalent. For instance, note that

$$\begin{aligned} \deg(\text{characteristic polynomial of } T) &= n_1 + \cdots + n_m \\ &= \dim G(\lambda_1, T) + \cdots + \dim G(\lambda_m, T) \\ &= \dim V, \end{aligned}$$

where the last equality follows from $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$. However, the minimal polynomial may have smaller degree, as shown in the following examples.

Example 22.8
Suppose $T \in \mathcal{L}(V)$ and the matrix of T is

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

It is clear that T has eigenvalue 2 with multiplicity 2, so the characteristic polynomial is $(x-2)^2$. However, it is also clear that $T - 2I = 0$, so the minimal polynomial of T is $x - 2$.

Example 22.9
Suppose $T \in \mathcal{L}(V)$ and the matrix of T is

$$\mathcal{M}(T) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

We see that this matrix is block diagonal with lower-triangular blocks:

$$\left(\begin{array}{cc|cc} 0 & 0 & & \\ 1 & 0 & & \\ \hline & & -1 & 0 \\ & & 0 & -1 \end{array}\right).$$

By Theorem 21.9[a], we can see that

$$V = G(0,T) \oplus G(-1,T),$$

where $G(0,T) = \operatorname{span}(e_1, e_2)$ and $G(-1,T) = \operatorname{span}(e_3, e_4)$. Thus, the characteristic polynomial of T is $p(x) = x^2(x+1)^2$.

The minimal polynomial must still include x and $x+1$ as factors, although they may have different exponents than in the characteristic polynomial. In other words, the minimal polynomial of T is of the form $m(x) = x^?(x+1)^?$, where the exponents are positive integers less than or equal to 2.

To find the minimal polynomial, we will lower each exponent as much as possible, while still maintaining the condition $m(T) = 0$. First, we will check if $T(T+1)^2 = 0$. By the matrix of T, we see that $Te_1 = e_2$ and $Te_2 = 0$.

$$T(T+1)^2 e_1 = (T+1)^2(Te_1) = (T+1)^2 e_2 = (T^2 + 2T + I)e_2 = e_2.$$

Thus, $T(T+1)^2 \neq 0$.

Now, we will check if $T^2(T+1) = 0$. Since $\mathcal{M}(T)$ is block-diagonal, we see that both $\operatorname{span}(e_1, e_2)$ and $\operatorname{span}(e_3, e_4)$ are T-invariant subspaces of V. It is easy to verify that $T^2|_{\operatorname{span}(e_1,e_2)} = 0$, so $T^2(T+1)|_{\operatorname{span}(e_1,e_2)} = 0$ because $T^2(T+1)$ is a multiple of T^2. Similarly, it is easy to see that $(T+1)|_{\operatorname{span}(e_3,e_4)} = 0$, so $T^2(T+1)|_{\operatorname{span}(e_3,e_4)} = 0$. Thus, $T^2(T+1) = 0$ over all V, so $T^2(T+1)$ is the mininal polynomial of T.

[a] Recall that Theorem 21.9 describes a block diagonal matrix with upper-triangular blocks, not lower-triangular. However, due to symmetry, essentially all results we have proved involving upper-triangular matrices also hold for lower-triangular matrices.

The following result gives some properties (without proof) of the minimal polynomial.

Fact 22.10
Suppose $T \in \mathcal{L}(V)$. Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T with multiplicities $n_1, \ldots, n_m$. Then,

- the minimal polynomial exists and is unique,
- the minimal polynomial is of the form $(x-\lambda_1)^{d_1} \cdots (x-\lambda_m)^{d_m}$ for $1 \leq d_i \leq n_i$.

Example 22.11
Recall Example 22.8. Consider a similar operator $T \in \mathcal{L}(V)$ such that

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

It is clear that the characteristic polynomial of T is $(x-2)^2$; thus, the minimal polynomial can be either $x-2$ or $(x-2)^2$. We can easily verify that $T-2I \neq 0$, so $x-2$ cannot be the minimal polynomial. Therefore, the minimal polynomial of T is also $(x-2)^2$.

In summary, the characteristic polynomial tells us the set of eigenvalues and the multiplicity of each eigenvalue for any operator $T \in \mathcal{L}(V)$. The minimal polynomial differs from the characteristic polynomial because it may have lower exponents than the characteristic polynomial.

22.4 Jordan Form

Now, we will introduce the Jordan form of an operator, which will allow us to write the matrix of an operator in a nice form using the material we have covered so far about eigenvalues and multiplicity.

First, to motivate the Jordan form, consider the following computational problem. Suppose you are given $\mathcal{M}(T)$ for some operator $T \in \mathcal{L}(V)$ and you wish to compute $T^9 + 2T^5 - T^3$. In general, if $\mathcal{M}(T)$ is a large matrix, computing large powers will require a very large number of calculations.

A way to make computing powers of large matrices faster is by finding an invertible operator S such that STS^{-1} is diagonal, since powers of diagonal matrices are very easy to compute. This works because

$$(STS^{-1})^n = (STS^{-1}) \cdots (STS^{-1}) = ST^nS^{-1},$$

so it is easy to recover T^n (hard to compute) from $(STS^{-1})^n$ (easy to compute).

However, as we have seen previously, not all operators are diagonalizable. But, the Jordan form of a matrix tries to solve this problem by finding a matrix representation which is close to diagonal for all operators.

We will start by investigating Jordan form for nilpotent operators.

Guiding Question
Suppose $N \in \mathcal{L}(V)$ is a nilpotent operator. Can we find a basis $v_1, \ldots, v_n$ of V such that $\mathcal{M}(N)$ is as simple as possible?

First, consider some small examples.

Example 22.12
Consider the case where $\dim V = 2$. The case where $N = 0$ is uninteresting because $\mathcal{M}(N)$ will always be the zero matrix. Thus, assume $N \neq 0$. We claim that there exists a basis v_1, v_2 of V such that

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

To prove this, note that $\dim \operatorname{Null}(N)$ must be between 0 and 2, inclusive. If $\dim \operatorname{Null}(N) = 2$, that would imply that $N = 0$, which we assumed not to be the case. Then, since N is nilpotent and $\dim V = 2$, it follows that $N^2 = 0$. This implies

$$N^2 v = N(Nv) = 0$$

for any $v \in V$, so $Nv \in \operatorname{Null}(N)$. Because $N \neq 0$, there exists some v such that $Nv \neq 0$, so $\dim \operatorname{Null}(N) > 0$. Therefore, $\dim \operatorname{Null}(N) = 1$.

Now, take $v_1 \notin \operatorname{Null}(N)$ and $v_2 = Nv_1$. This implies that $v_2 = Nv_1 \in \operatorname{Null}(N)$, so v_1, v_2 is linearly independent. Thus, v_1, v_2 forms a basis of V. Then, because $Nv_2 = 0$ and $Nv_1 = v_2$, it follows that

$$\mathcal{M}(T, (v_2, v_1)) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Example 22.13
Consider the case where $\dim V = 3$. If $N = 0$, then $\mathcal{M}(N)$ is always the zero matrix:

$$\mathcal{M}(N) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Now, assume $N \neq 0$. By similar logic to the previous example, we can show that $\dim \operatorname{Null}(N)$ cannot be equal to 0 nor 3. Thus, we have two cases.

If $\dim \operatorname{Null}(N) = 2$, then there exists a basis such that

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

On the other hand, if $\dim \operatorname{Null}(N) = 1$, then there exists a basis such that

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

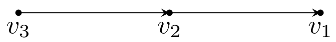
Let us further investigate the matrices in the above example. Suppose v_1, v_2, v_3 is a basis of V such that

$$\mathcal{M}(N, (v_1, v_2, v_3)) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

From this matrix, we can determine that

$$\begin{aligned} Nv_1 &= 0, \\ Nv_2 &= v_1, \\ Nv_3 &= v_2. \end{aligned}$$

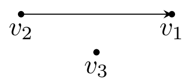
We can represent this operator with the following diagram



where each dot represents a basis vector and each arrow represents applying N to each vector. With this diagram notation, we can also describe the scenario where

$$\mathcal{M}(N, (v_1, v_2, v_3)) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

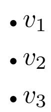
with the diagram



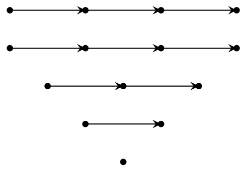
Finally, we can describe the scenario where

$$\mathcal{M}(N, (v_1, v_2, v_3)) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

with the diagram



Now, we will use the diagram notation to expand our discussion to Jordan form of nilpotent operators in vector spaces of any dimension. In the Jordan form, all basis vectors should be included in one of these "chains":



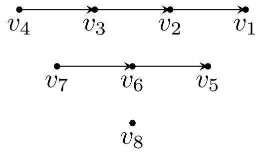
Note that there can be any number of these chains; the only requirement is that the total number of dots should be $\dim V$. These diagrams uniquely determine a matrix based on the length of each chain; for instance, the above diagram corresponds to the block-diagonal matrix

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

23 Jordan Form (continued), Trace, and Determinant

23.1 Review

Last time, we discussed how for any nilpotent operator $N \in \mathcal{L}(V)$, there exists a basis $v_1, \ldots, v_n$ such that N can be expressed with diagrams of the following form:



In terms of matrices, $\mathcal{M}(N)$ with respect to this basis would look like

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

23.2 Jordan Form (continued)

The above matrix is in Jordan form (also known as Jordan canonical form). Now, we are ready to give a definition of the Jordan form.

Definition 23.1 (Jordan block and Jordan form)
A **Jordan block** is a matrix of the form

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

A **Jordan form** is a block-diagonal matrix where all blocks are Jordan blocks.

Note that for a matrix a Jordan form, it not needed that all Jordan blocks have the same λ along the diagonal. For instance, the matrix

$$\begin{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} & & \\ & \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} & \\ & & (2) \end{pmatrix}$$

is a Jordan form. The Jordan blocks are highlighted with parentheses.

Theorem 23.2
Suppose V is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Then,

1. there exists a basis under which $\mathcal{M}(T)$ is a Jordan form,
2. the Jordan blocks that appear in $\mathcal{M}(T)$ are uniquely determined by T up to permutation.

Proof. We will only prove the existence property of the above theorem. In the previous lecture, we gave some intuition (although not a rigorous proof) on why this statement holds if T is nilpotent. To extend to all T, recall that

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T).$$

By definition, $(T - \lambda_i I)|_{G(\lambda_i, T)}$ is a nilpotent operator. It follows that there exists a basis of $G(\lambda_i, T)$ such that $\mathcal{M}((T - \lambda_i I)|_{G(\lambda_i, T)})$ is a Jordan form. We know that

$$\mathcal{M}(T|_{G(\lambda, T)}) = \mathcal{M}((T - \lambda_i I)|_{G(\lambda_i, T)}) + \mathcal{M}(\lambda_i I),$$

so $\mathcal{M}(T|_{G(\lambda, T)})$ is also a Jordan form because $\mathcal{M}(\lambda_i I)$ only adds values along the diagonal. Putting all such bases together forms a basis of v such that $\mathcal{M}(T)$ is a Jordan form. □

In the last lecture, we discussed how Jordan form would help us compute large powers of operators quickly.

Example 23.3
Suppose $T \in \mathcal{L}(V)$ and we wish to compute T^{100}. To do this, we know by Theorem 23.2 that there exists a basis of V such that $\mathcal{M}(T)$ is in Jordan form. For instance,

$$\mathcal{M}(T) = \begin{pmatrix} \begin{pmatrix} \lambda_1 & 1 & \\ & \lambda_1 & 1 \\ & & \lambda_1 \end{pmatrix} & & \\ & \begin{pmatrix} \lambda_2 & 1 \\ & \lambda_2 \end{pmatrix} & \\ & & \begin{pmatrix} \lambda_3 \end{pmatrix} \end{pmatrix}.$$

To compute the 100^{th} power of a matirx, we simply take the 100^{th} power of each of its blocks. To compute the power of each block, we can express each block as $\lambda I + J$, where

$$J = \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}.$$

By the binomial theorem,

$$(\lambda I + J) = (\lambda I)^{100} + \binom{100}{1}(\lambda I)^{99} J + \binom{100}{2}(\lambda I)^{98} J^2 + \cdots + J^{100}.$$

It is clear that $(\lambda I)^i = \lambda^i I$. Furthermore, it can be deduced that

$$J^m = \begin{pmatrix} 0 & \cdots & 1 & & \\ & \ddots & & \ddots & \\ & & \ddots & & 1 \\ & & & \ddots & \vdots \\ & & & & 0 \end{pmatrix},$$

where the 1 in the first row is in the $(m+1)^{\text{th}}$ column. Thus, using the 5×5 case as an example, we can calculate

$$(\lambda I + J)^{100} = \begin{pmatrix} \lambda^{100} & \binom{100}{1}\lambda^{99} & \binom{100}{2}\lambda^{98} & \binom{100}{3}\lambda^{97} & \binom{100}{4}\lambda^{96} \\ & \lambda^{100} & \binom{100}{1}\lambda^{99} & \binom{100}{2}\lambda^{98} & \binom{100}{3}\lambda^{97} \\ & & \lambda^{100} & \binom{100}{1}\lambda^{99} & \binom{100}{2}\lambda^{98} \\ & & & \lambda^{100} & \binom{100}{1}\lambda^{99} \\ & & & & \lambda^{100} \end{pmatrix}.$$

Note that if the exponent is smaller than the size of the matrix, the upper-right corner of the matrix would be filled with zeros. By applying this logic to each block of $\mathcal{M}(T)$, we have a way to quickly compute $\mathcal{M}(T^{100})$, as desired.

23.3 Trace

We will now move on to the final chapter of the textbook, first with our discussion of trace.

Definition 23.4 (trace)

1. The **trace** of a square matrix A is defined to be the sum of the diagonal elements of A.
2. The **trace** of an operator $T \in \mathcal{L}(V)$ is defined to be the sum of the eigenvalues of T, counted with multiplicity.

At first glance, these two definitions seem completely unrelated. For now, we will use $\operatorname{tr}(A)$ to denote the trace of a matrix and $\operatorname{Tr}(T)$ to denote the trace of an operator. Later, we will show that these definitions are actually equivalent, after which we will use tr for both cases.

First, we will study some properties of the trace of a matrix.

Theorem 23.5 (Properties of trace of a matrix)
Suppose A and B are square matrices of the same size. Then,

1. $\operatorname{tr}(A + B) = \operatorname{tr}(A) + \operatorname{tr}(B)$,
2. $\operatorname{tr}(\lambda A) = \lambda \operatorname{tr}(A)$.

These properties are easy to verify by matrix addition and scalar multiplication. Note that the above properties show that $\operatorname{tr} : \mathbb{C}^{n,n} \to \mathbb{C}$ is a linear map.

Lemma 23.6
Suppose A and B are square matrices of the same size. Then, $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

Proof. By matrix multiplication, we can calculate

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

Thus,

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

Similarly, we can deduce

$$\operatorname{tr}(BA) = \sum_{i=1}^{n} \sum_{j=1}^{n} B_{i,j} A_{j,i}.$$

By switching the variable names i and j and switching the order of the summations, it follows that

$$\begin{aligned}
\operatorname{tr}(AB) &= \sum_{i=1}^{n} \sum_{j=1}^{n} A_{i,j} B_{j,i} \\
&= \sum_{j=1}^{n} \sum_{i=1}^{n} A_{j,i} B_{i,j} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} B_{i,j} A_{j,i} \\
&= \operatorname{tr}(BA),
\end{aligned}$$

as desired. □

Now, suppose $T \in \mathcal{L}(V)$ and $v_1, \ldots, v_n$ and $u_1, \ldots, u_n$ are bases of V. Let $A = \mathcal{M}(T, (v_1, \ldots, v_n))$ and $B = \mathcal{M}(T, (u_1, \ldots, u_m))$. Recall in an earlier lecture we showed that there exists a change of basis matrix S and $B = S^{-1}AS$.

We are now ready to prove the two definitions of trace are equivalent.

Theorem 23.7
Suppose $T \in \mathcal{L}(V)$. Then, $\text{Tr}(T) = \text{tr}(\mathcal{M}(T))$.

Proof. First, we will show that $\text{tr}(\mathcal{M}(T))$ is independent of the choice of basis. Let $v_1, \ldots, v_n$ and $u_1, \ldots, u_n$ be bases of V and let $A = \mathcal{M}(T, (v_1, \ldots, v_n))$ and $B = \mathcal{M}(T, (u_1, \ldots, u_n))$. If S is the change of basis matrix, then $B = S^{-1}AS$. It follows that

$$\begin{aligned}\text{tr}(B) &= \text{tr}(S^{-1}AS)\\ &= \text{tr}((AS)S^{-1})\\ &= \text{tr}(A),\end{aligned}$$

where the second equality follows from Lemma 23.6.

Now, since we have shown that $\mathcal{M}(T)$ is independent of the chosen basis, we only need to show that $\text{Tr}(T) = \text{tr}(\mathcal{M}(T))$ for some basis of V. Let $v_1, \ldots, v_n$ be a basis such that $A = \mathcal{M}(T, (v_1, \ldots, v_n))$ is upper-triangular. Then, the diagonal entries of A are equivalent to the eigenvalues of T with multiplicity. Therefore, $\text{Tr}(T) = \text{tr}(A)$, as desired. □

This theorem is quite fascinating: the left-hand side only depends on the operator T, while the right-hand side depends on both T and an arbitrary basis $v_1, \ldots, v_n$. Additionally, why is the trace of a matrix defined as the sum of its diagonal entries? Can we instead define the trace as another function of the matrix (e.g. the sum of the entries on the other diagonal or the sum of the entries of any of the columns/rows) and still have Theorem 23.7 hold? The answer turns out to be no, which shows the beauty of the definition of trace.

Example 23.8
Suppose $T \in \mathcal{L}(V)$ with matrix representation

$$\mathcal{M}(T) = \begin{pmatrix} 1 & 3 & 2 \\ 4 & 0 & 5 \\ 7 & 9 & -1 \end{pmatrix}.$$

At first glance, it is impossible to know what the eigenvalues of T are. However, Theorem 23.6 tells us that the sum of the eigenvalues is equal to $\text{tr}(\mathcal{M}(T)) = 1 + 0 + (-1) = 0$. Therefore, if we know two of the eigenvalues of T, then we are able to easily find the third.

Furthermore, recall from Definition 22.1 that the characteristic polynomial of T is

$$(x - \lambda_1)^{n_1} \cdots (x - \lambda_m)^{n_m},$$

where $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues of T with multiplicities $n_1, \ldots, n_m$. Expanding the polynomial above, we find that the characteristic polynomial can be expressed as

$$x^n - (n_1\lambda_1 + \cdots + n_m\lambda_m)x^{n-1} + \cdots + (-1)^n(\lambda_1^{n_1} \cdots \lambda_m^{n_m}).$$

Thus, $\text{Tr}(T)$ equals the negative of the coefficient of the x^{n-1} term.

23.4 Determinant

Now, we will introduce the notion of determinant.

Definition 23.9 (determinant)

1. The **determinant** of an operator $T \in \mathcal{L}(V)$ is defined as the product of the eigenvalues of T, counted with multiplicity.
2. The **determinant** of a $n \times n$ square matrix A is defined by

$$\det(A) = \sum_{\sigma \in \text{perm}(n)} \text{sign}(\sigma) A_{1,\sigma(1)} \cdots A_{n,\sigma(n)}.$$

Once again, we will for now use $\text{Det}(T)$ to denote the determinant of an operator and $\det(A)$ to denote the determinant of an matrix.

Recall that the characteristic polynomial of an operator $T \in \mathcal{L}(V)$ can be expressed as

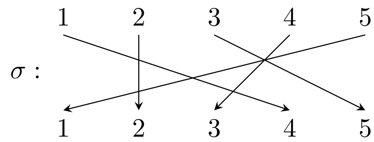
$$x^n - (n_1\lambda_1 + \cdots + n_m\lambda_m)x^{n-1} + \cdots + (-1)^n(\lambda_1^{n_1} \cdots \lambda_m^{n_m}).$$

Thus, it follows that $\text{Det}(T)$ equals $(-1)^n$ multiplied by the constant term of the characteristic polynomial.

Now, we will explain the formula for the determinant of a matrix:

$$\det(A) = \sum_{\sigma \in \text{perm}(n)} \text{sign}(\sigma) A_{1,\sigma(1)} \cdots A_{n,\sigma(n)}.$$

First, note that this summation has $n!$ terms, each of which is a product of n matrix entries. The notation $\text{perm}(n)$ represents the set of all functions that map $(1, 2, \ldots, n)$ to a permutation of $(1, 2, \ldots, n)$. For instance, we can visualize one such function σ with the following diagram:



In this scenario, $\sigma(1) = 4, \sigma(2) = 2$, and so on. Finally, if we draw such a diagram for arbitrary σ, then $\text{sign}(\sigma)$ is defined to be 1 if the number of crossings (i.e. the total number of times two arrows cross) is even and -1 if the number is odd.[39] For instance, the number of crossings in the above diagram is 7, so $\text{sign}(\sigma) = -1$.

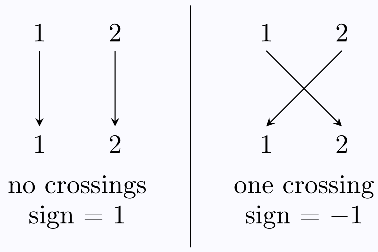
Consider some simple examples of determinants of matrices.

Example 23.10

Suppose we wish to find the determinant of the 2×2 matrix

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

We find that $\text{perm}(2)$ only has 2 elements:



Thus,

$$\det(A) = A_{11}A_{22} - A_{12}A_{21},$$

which is our familiar formula for the determinant of a 2×2 matrix.

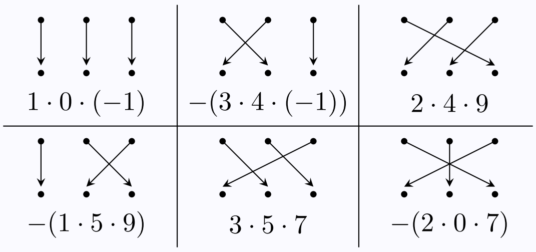
[39] A more formal (but equivalent) definition is that $\text{sign}(\sigma)$ is defined to be 1 if the number of pairs of integers (j, k) with $1 \le j < k \le n$ such that $\sigma(j)$ appears after $\sigma(k)$ in the list $(\sigma(1), \ldots, \sigma(n))$ is even and -1 if the number of such pairs is odd.

Example 23.11

Suppose we wish to find the determinant of the 3×3 matrix

$$A = \begin{pmatrix} 1 & 3 & 2 \\ 4 & 0 & 5 \\ 7 & 9 & -1 \end{pmatrix}.$$

There are 6 permutations:



Thus,

$$\det(A) = (1 \cdot 0 \cdot (-1)) - (3 \cdot 4 \cdot (-1)) + (2 \cdot 4 \cdot 9) - (1 \cdot 5 \cdot 9) + (3 \cdot 5 \cdot 7) - (2 \cdot 0 \cdot 7) = 144.$$

Now, similar to what we did with trace, we will show that the two definitions of determinant are equivalent.

Theorem 23.12

Suppose $T \in \mathcal{L}(V)$. Then, $\mathrm{Det}(T) = \det(\mathcal{M}(T))$.

We will cover the proof of this result in the next lecture.

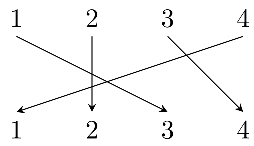
24 Determinant (continued)

24.1 Review

Last time, we introduced the definition for the determinant of an operator and the determinant of a matrix (see Definition 23.9). In particular, recall the formula for the determinant of a matrix A:

$$\det(A) = \sum_{\sigma \in \text{perm}(n)} \text{sign}(\sigma) A_{1,\sigma(1)} \cdots A_{n,\sigma(n)}.$$

In other words, each entry in the summation is the product of n matrix entries, one from each row, such that no two entries are in the same column. For instance, the permutation

σ :


would correspond to the product of the matrix elements highlighted below:

$$\begin{pmatrix} & & A_{13} & \\ & A_{22} & & \\ & & & A_{34} \\ A_{41} & & & \end{pmatrix}.$$

Additionally, recall that

$$\text{sign}(\sigma) = (-1)^{\text{parity of number of crossings in diagram of } \sigma}.$$

At the end of last lecture, we introduced Theorem 23.12, which relates the two definitions of determinant. We will prove this result later in this lecture.

24.2 Determinant (continued)

First, we will discuss some properties of determinants.

Theorem 24.1 (Properties of $\det(A)$)
Suppose A is a square matrix. Then,

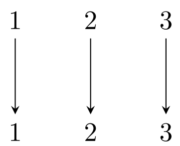
0. multiplying a row/column by some scalar c multiplies $\det(A)$ by c,
1. interchanging two rows/columns of A flips the sign of $\det(A)$,
2. if A has two rows/columns that are equivalent, then $\det(A) = 0$,
3. $\det(A)$ is additive in rows/columns,
4. adding $c \cdot (i^{\text{th}}$ row) to the j^{th} row does not change $\det(A)$

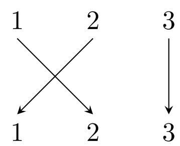
Proof. To prove (0), note that each permutation contains one entry from each row and column of the matrix. Thus, multiplying a row/column by c multiplies each summand by c, so the determinant is multiplied by c.

To prove (1), let A' be the result when we switch the first two rows of the following 3×3 matrix A:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \; A' = \begin{pmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

Also, let σ and τ be the permutations

σ :


τ :


Note that σ for matrix A and τ for matrix A' both correspond to the product $a_{11}a_{22}a_{33}$. However, $\operatorname{sign}(\sigma) = -\operatorname{sign}(\tau)$, so σ contributes $a_{11}a_{22}a_{33}$ to $\det(A)$ while τ contributes $-a_{11}a_{22}a_{33}$ to $\det(A')$. It can be shown that this logic extends to all such permutations and can be generalized to matrices of arbitrary size, so $\det(A) = -\det(A')$. The proof that the sign of $\det(A)$ is flipped when two columns are interchanged is similar.

To prove (2), note that switching the two identical rows/columns gives the same matrix A. Thus, (1) implies that $\det(A) = -\det(A)$, so $\det(A) = 0$. Furthermore, note that (0) also implies that if two rows/columns of A are scalar multiples of each other, then $\det(A) = 0$.

To prove (3), suppose we have two matrices A and B of the same size which are identical except for a single row:

$$A = \begin{pmatrix} \vdots & \cdots & \vdots \\ a_1 & \cdots & a_n \\ \vdots & \cdots & \vdots \end{pmatrix}, B = \begin{pmatrix} \vdots & \cdots & \vdots \\ b_1 & \cdots & b_n \\ \vdots & \cdots & \vdots \end{pmatrix}.$$

Let matrix C also be identical to both A and B except for a single row:

$$C = \begin{pmatrix} \vdots & \cdots & \vdots \\ a_1 + b_1 & \cdots & a_n + b_n \\ \vdots & \cdots & \vdots \end{pmatrix}.$$

Let the different row be the i^{th} row. Then,

$$C_{i,\sigma(i)} = A_{i,\sigma(i)} + B_{i,\sigma(i)}$$

for all permutations σ. By the definition of determinant of a matrix, it follows that $\det(C) = \det(A) + \det(B)$.

To prove (4), let B be the matrix that is identical to A except the j^{th} row of B is $c \cdot (i^{\text{th}}$ row) of A. By (2), $\det(B) = 0$ because the i^{th} and j^{th} rows are scalar multiples of each other. Then, by (3), $\det(A + B) = \det(A) + \det(B) = \det(A)$, as desired. □

The next result will show that determinant is multiplicative.

Proposition 24.2
Suppose A and B are square matrices of the same size. Then, $\det(AB) = \det(A)\det(B)$.

Proof. Recall that we can perform Gaussian elimination on A to obtain a matrix in row-reduced echelon form. Recall the three elementary row operations:

$$\begin{aligned} S(i,j) &= \text{swap rows } i \text{ and } j, \\ M(i;c) &= \text{multiply the } i^{\text{th}} \text{ row by } c, \\ A(i \xrightarrow{c} j) &= \text{add } c \text{ times the } i^{\text{th}} \text{ row to the } j^{\text{th}} \text{ row.} \end{aligned}$$

Now, note that performing any elementary row operation on A is equivalent to left-multiplying A by some matrix.[40] Let A' be the result of performing some elementary row operation e on matrix A and let E be the matrix corresponding to e. It follows that $A' = EA$. Additionally, performing e on matrix AB gives $E(AB) = (EA)B = A'B$.

Suppose $e = S(i,j)$. By Theorem 24.1, it follows that $\det(A') = -\det(A)$ and $\det(A'B) = -\det(AB)$. In this scenario, note that proving $\det(A'B) = \det(A')\det(B)$ will directly imply $\det(AB) = \det(A)\det(B)$. Similarly, it can be seen for the other two elementary row operations that proving $\det(A'B) = \det(A')\det(B)$ is sufficient.

Thus, we can continue performing elementary row operations on A until we obtain a row-reduced echelon form R. By the same logic as above, it is sufficient to show that $\det(RB) = \det(R)\det(B)$. If R has a row of zeros, then it can be seen by Theorem 24.1 or the formula for determinant that $\det(R) = 0$. Additionally, RB will

[40] You can learn more about this here: https://en.wikipedia.org/wiki/Elementary_matrix.

also have a row of zeros, so $\det(RB) = \det(R)\det(B) = 0$. If R does not have a row of zeros, then because R is square, it must take the form

$$R = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} = I.$$

Since $\det(I) = 1$,

$$\det(RB) = \det(IB) = \det(B) = \det(I)\det(B) = \det(R)\det(B),$$

as desired. □

We are now ready to prove our desired theorem.

Theorem 24.3
Suppose $T \in \mathcal{L}(V)$. Then, $\mathrm{Det}(T) = \det(\mathcal{M}(T))$.

Proof. Choose a basis $v_1, \dots, v_n$ of V such that $\mathcal{M}(T, (v_1, \dots, v_n))$ is upper-triangular. Let $A = \mathcal{M}(T, (v_1, \dots, v_n))$ and let the diagonal elements of A be $a_1, \dots, a_n$:

$$\begin{pmatrix} a_1 & & * \\ & \ddots & \\ 0 & & a_n \end{pmatrix}.$$

The eigenvalues of T are the diagonal elements, so $\det(T) = a_1 \cdots a_n$. Furthermore, it can be seen that if σ does not map every element to itself, then at least one of $A_{1,\sigma(1)}, \dots, A_{n,\sigma(n)}$ must be below the diagonal and thus equal to 0. Therefore,

$$\begin{aligned} \det(A) &= \sum_{\sigma \in \mathrm{perm}(n)} \mathrm{sign}(\sigma) A_{1,\sigma(1)} \cdots A_{n,\sigma(n)} \\ &= 1 \cdot A_{11} \cdots A_{nn} \\ &= a_1 \cdots a_n \end{aligned}$$

, so $\mathrm{Det}(T) = \det(A)$. Now, let $u_1, \dots, u_n$ be an arbitrary basis of V and let $B = \mathcal{M}(T, (u_1, \dots, u_n))$. Let S be the change of basis matrix, so $B = S^{-1}AS$. Then,

$$\begin{aligned} \det(B) &= \det(S^{-1}AS) \\ &= \det(S^{-1})\det(A)\det(S) \\ &= (\det(S^{-1})\det(S))\det(A) \\ &= \det(S^{-1}S)\det(A) \\ &= \det(A), \end{aligned}$$

where the second and fourth equalities follow from Proposition 24.2. Therefore, $\mathrm{Det}(T) = \det(\mathcal{M}(T))$ under any basis of V. □

From now on, we will use det as notation for the determinant of both operators and matrices.

24.3 Finding Eigenvalues and Eigenvectors

The next result will relate the determinant and the characteristic polynomial.

Proposition 24.4
Suppose $T \in \mathcal{L}(V)$. Then, the characteristic polynomial of T equals $p_T(x) = \det(xI - T)$.

Proof. Choose a basis of V such that $\mathcal{M}(T)$ is upper-triangular. Then, the diagonal entries of $\mathcal{M}(T)$ are the eigenvalues $\lambda_1, \dots, \lambda_n$. It follows that $\mathcal{M}(xI - T)$ has diagonal entries $x - \lambda_1, \dots, x - \lambda_n$. Therefore,

$$\det(\mathcal{M}(xI - T)) = (x - \lambda_1) \cdots (x - \lambda_n),$$

as desired. □

The above result finally gives us a way to compute eigenvalues of T. Let $A = \mathcal{M}(T)$ under any basis. Then, $\det(xI - A)$ gives a monic polynomial of degree n. Solving for the roots (with multiplicity) of this polynomial gives the eigenvalues of T.

Example 24.5
Suppose V is a two-dimensional vector space and $T \in \mathcal{L}(V)$. Then, we know that the characteristic polynomial of T is

$$p_T(x) = (x - \lambda_1)(x - \lambda_2) = x^2 - (\lambda_1 + \lambda_2) + \lambda_1\lambda_2 = x^2 - (\operatorname{tr} T)x + \det T.$$

Let $\mathcal{M}(T) = A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then, $\operatorname{tr} T = \operatorname{tr} A = a + d$ and $\det T = \det A = ad - bc$, so

$$p_T(x) = x^2 - (a + d)x + (ad - bc),$$

which is a formula we have used before to compute the eigenvalues of a 2×2 matrix.

Now, we will discuss how to compute eigenvectors.

Example 24.6
Suppose $T \in \mathcal{L}(V)$ and $A = \mathcal{M}(T)$ is the upper-triangular matrix

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{pmatrix}.$$

Because A is upper-triangular, the eigenvalues are $1, 2, 3$. Now, we will explore some examples of finding eigenvalues and eigenvectors.

1. For $\lambda = 1$, it is clear that e_1 is an eigenvector.
2. For $\lambda = 2$, an eigenvector must be in $\operatorname{Null}(A - 2I)$. We can compute

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

   Performing Gaussian elimination on this matrix gives

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

   From this matrix, it is clear that $\operatorname{Null}(A - 2I)$ is spanned by $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$.

3. For $\lambda = 3$, we can follow the same process as for $\lambda = 2$. In general, you can follow the same process as above for any matrix/eigenvalue.

Example 24.7

Suppose $T \in \mathcal{L}(V)$ and $A = \mathcal{M}(T)$ is the block-diagonal matrix

$$A = \left(\begin{array}{ccc|cc} 1 & 8 & 6 & 0 & 0 \\ 2 & 9 & 7 & 0 & 0 \\ 3 & 4 & 5 & 0 & 0 \\ \hline 0 & 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 0 & 2 \end{array}\right).$$

Note that the upper-left block is relatively complicated and would require more computation to compute eigenvalues. However, $U = \operatorname{span}(e_1, e_2, e_3)$ is a T-invariant subspace. Then, by Lemma 12.13, -1 and 2 are eigenvalues of T because they are eigenvalues of the lower-right block of A.

Example 24.8

Suppose $T \in \mathcal{L}(V)$ and $A = \mathcal{M}(T)$ is the block upper-triangular matrix

$$A = \left(\begin{array}{cc|cc} 1 & 2 & 2 & 3 \\ 3 & 4 & -1 & 1 \\ \hline 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 2 \end{array}\right).$$

We wish to find all Jordan blocks with diagonal entry equal to 2 (after expressing A in Jordan form). Since A only has two 2's on the diagonal, there are either two 1×1 Jordan blocks or one 2×2 Jordan block with 2 on the diagonal.

To determine this, we compare $E(2, T)$ and $G(2, T)$. If $E(2, T) = G(2, T)$, then there are two 1×1 Jordan blocks. If $E(2, T) \neq G(2, T)$, then there is one 2×2 Jordan blocks. This simply requires computing $\dim E(2, T)$, which can be done using Gaussian elimination.