MATH 110 Lecture Notes

Alec Li

Fall 2021 — Professor Zvezdelina Stankova

Contents

Theorems

Definitions

8/25/2021

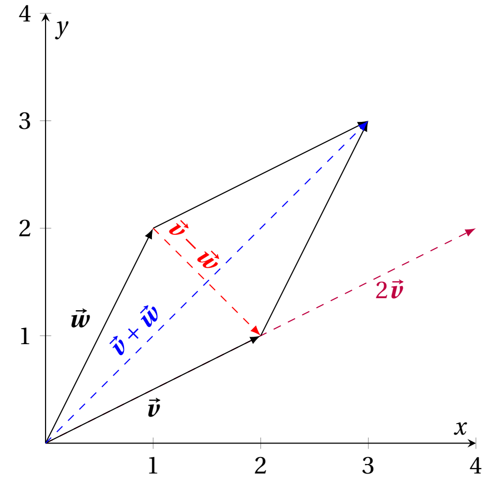
Lecture 1

Vector Spaces and Fields

1.1 Vectors

You should be familiar with $\mathbb{R}^2$—the real plane. $\mathbb{R}^2$ is the prototype for us, and we will generalize these concepts for all types of vector spaces.

Suppose we have $\vec{\boldsymbol{v}} = (v_1, v_2)$ and $\vec{\boldsymbol{w}} = (w_1, w_2)$.



This parallelogram construction is the typical illustration of vector addition and subtraction:

To add or subtract two vectors, we add/subtract element-wise: that is, $\vec{\boldsymbol{v}} \pm \vec{\boldsymbol{w}} = (v_1 \pm w_1, v_2 \pm w_2)$.

To multiply a vector by a scalar, we multiply each element by the scalar: $a\vec{\boldsymbol{v}} = (av_1, av_2)$, where $a \in \mathbb{R}$.

If we were to generalize these concepts, we can extend to an n-dimensional *vector space*. Any vector in this vector space can be written as

$$\vec{\boldsymbol{v}} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}.$$

Each a_i are objects in the vector space–we can expand the scope of these objects to include things other than numbers.

1.1.1 Operations

We have two operations: addition and scalar multiplication, as mentioned before.

What does it mean to have operations?

- Addition: $(\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}}) \mapsto \vec{\boldsymbol{x}} + \vec{\boldsymbol{y}} \in \mathbb{R}^n$
- Scalar multiplication: $(a, \vec{\boldsymbol{x}}) \mapsto \vec{\boldsymbol{a}}\boldsymbol{x} \in \mathbb{R}^n$; here, $a \in \mathbb{R}$

Here, $\mathbb{R}$ is called a "field"; it's a vector space with some more additional properties.

1.2 Vector Space Axioms

We will split this into two parts. The first part relates to addition.

1. Commutativity (comm.): $\vec{x}+\vec{y}=\vec{y}+\vec{x}$
2. Associativity (ass.): $(\vec{x}+\vec{y})+\vec{z}=\vec{x}+(\vec{y}+\vec{z})$
3. Additive Identity (add. iden.): we have a $\vec{0}$; $\vec{0}+\vec{x}=\vec{x}+\vec{0}=\vec{x}$
4. Additive Inverses (add. inv.): $(\forall\vec{x}\in\mathbb{R}^n)(\exists\vec{y}\in\mathbb{R}^n)\big(\vec{x}+\vec{y}=\vec{0}\big)$.

The second part relates to scalar multiplication.

1. Multiplicative identity $1\in\mathbb{R}$: $1\cdot\vec{x}=\vec{x}$
2. Associativity: $(ab)\vec{x}=a(b\vec{x})$; here, $a,b\in\mathbb{R}$.
3. Distributive Laws:
   (a) $(a+b)\vec{x}=a\vec{x}+b\vec{x}$
   (b) $a(\vec{x}+\vec{y})=a\vec{x}+a\vec{y}$

Further, we must have that the vector space $\mathcal{V}$ is a non-empty set—that is, $\mathcal{V}\neq\varnothing$.

In abstracting this, we will keep all essential information, and forget all other information. For example, we will forget that we are working with numbers, and instead we will be working with arbitrary objects.

Definition 1.1: Set of Functions

For any two sets S and T, $\mathcal{F}(S,T)$ is the set of functions $f: S\to T$. When $T=F$ is a field, then $\mathcal{F}(S,F)$ is a vector space over F, with addition and scalar multiplication defined by

$$(f+g)(s)=f(s)+g(s) \qquad (\lambda\cdot f)(s)=\lambda f(s)$$

Example 1.2

We have $\mathcal{F}(\mathbb{R},\mathbb{R})$ = set of all functions $f:\mathbb{R}\to\mathbb{R}$.

Here, we will be specifically looking at all quadratic functions; that is, all $f(x)=ax^2+bx+c$, where $a,b,c\in\mathbb{R}$. We will denote this space as $P_2=\{\text{polynomials}/\mathbb{R}\mid\text{degree}\leq 2\}$. That is, the set of all polynomials over the reals with degree at most 2.

The vectors in the vector space are our functions $f(x)$, which we can think of as $\vec{f}$.

To show that p_2 is a vector space (at this point), we must show that the definition of vector space holds. That is, all 8 properties discussed above must hold, along with the fact that the vector space is closed: $f,g\in P_2 \implies f+g\in P_2$, $c\in\mathbb{R}\implies cf\in P_2$

One thing to notice here is that $P_2\cong\mathbb{R}^3$; that is P_2 is isomorphic to $\mathbb{R}^3$; we can represent each polynomial as the vector $\begin{bmatrix}a\\b\\c\end{bmatrix}$.

Example 1.3

We can also look at matrices; for example, $\begin{bmatrix}a & b\\c & d\end{bmatrix}$.

The space of all 2×2 matrices on reals is denoted as $M_{2\times2}(\mathbb{R})$.

Addition and scalar multiplication is defined as

$$\lambda \begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix} + \begin{bmatrix} a_2 & b_2 \\ c_2 & d_2 \end{bmatrix} = \begin{bmatrix} \lambda a_1 + a_2 & \lambda b_1 + b_2 \\ \lambda c_1 + c_2 & \lambda d_1 + d_2 \end{bmatrix}.$$

We can also note that $M_{2\times 2}(\mathbb{R}) \cong \mathbb{R}^4$; each matrix can be represented as a column vector $\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$.

Example 1.4

We can also look at differential equations, for example, $f'' + f = 0$.

The set of all solutions $\mathscr{D} = \{f : \mathbb{R} \to \mathbb{R} \mid f'' + f = 0\}$.

Some possible solutions are $f_1 = \sin x$, $f_2 = \cos x$. If $\mathscr{D}$ is a vector space, then $f = a \sin x + b \cos x$ is also a solution, where $a, b\ in \mathbb{R}$. This construction is also called a *linear combination*.

It turns out that $\mathscr{D}$ is in fact a vector space, and isomorphic to $\mathbb{R}^2$, with the form of f as above.

Here is a formal definition of a vector space:

Definition 1.5: Vector Space

A *vector space* over a field F (see below) is a set $\mathcal{V}$ of vectors along with two operations:

- *Addition*: $\vec{x}, \vec{y} \mapsto \vec{x} + \vec{y}$, where $\vec{x}, \vec{y} \in \mathcal{V}$
- *Scalar multiplication*: $\vec{x}, \vec{y} \mapsto \lambda \cdot \vec{x}$, where $\vec{x} \in \mathcal{V}$, $\lambda \in F$

Addition must have the following properties for all $\vec{x}, \vec{y}, \vec{z} \in \mathcal{V}$ and $a, b \in \mathbb{R}$:

1. *Commutativity*: $\vec{x} + \vec{y} = \vec{y} + \vec{x}$
2. *Associativity*: $(\vec{x} + \vec{y}) + \vec{z} = \vec{x} + (\vec{y} + \vec{z})$
3. There is a *zero vector* $\vec{0}$ with $\vec{0} + \vec{x} = \vec{x} + \vec{0} = \vec{x}$
4. There are *inverses*: for every $\vec{x} \in \mathcal{V}$ there is a $\vec{y} \in \mathcal{V}$ with $\vec{x} + \vec{y} = \vec{0}$.

Scalar multiplication must also satisfy:

5. *Identity*: $1 \cdot \vec{x} = \vec{x}$
6. *Associativity*: $(ab) \cdot \vec{x} = a \cdot (b\vec{x})$
7. *Distributivity I*: $(a + b)\vec{x} = a\vec{x} + b\vec{x}$
8. *Distributivity II*: $a\left(\vec{x} + \vec{y}\right) = a\vec{x} + a\vec{y}$

1.3 Fields

Vector spaces are defined over fields. What are these fields?

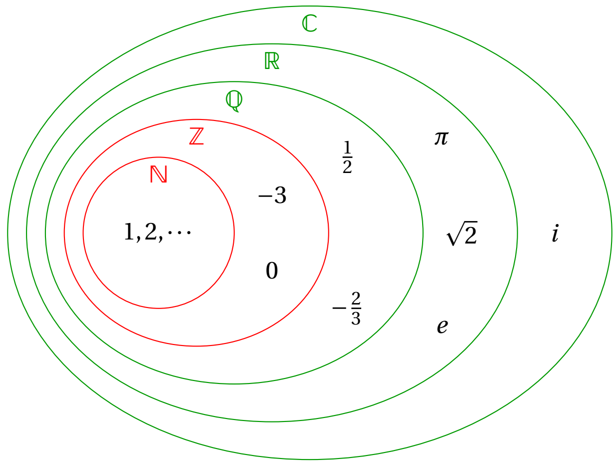


Figure 1: Five famous sets

The naturals ($\mathbb{N}$) and the integers ($\mathbb{Z}$) are not fields, but the rationals ($\mathbb{Q}$), the reals ($\mathbb{R}$), and the complexes ($\mathbb{C}$) are fields.

Fields must be closed under addition, subtraction, multiplication, and division. They must also be commutative with addition and multiplication, and have units and inverses for both operations. That means that there also needs to be a multiplicative inverse (something that vector spaces do not have): $\frac{1}{a} \in F$, where $a \neq 0$.

Our vector spaces will primarily be taking scalars from $\mathbb{R}$ and $\mathbb{C}$, but especially $\mathbb{R}$. We will also be working with fields like $\mathbb{Z}_2 = \{0, 1\}$. That is, the integers modulo 2.

One question to ask is: what is the difference between a vector space over the reals ($V_{\mathbb{R}}$) and a vector space over the complex numbers ($V_{\mathbb{C}}$)?

Example 1.6

Consider the vector space of complex numbers over complex numbers: $\mathbb{C}_{\mathbb{C}} = \{c \cdot 1 \mid c \in \mathbb{C}\}$. This vector space only depends on the complex number we choose, so it looks like $\mathbb{C}$.

Consider the vector space of complex numbers over reals: $\mathbb{C}_{\mathbb{R}} = \{a \cdot 1 + b \cdot i \mid a, b \in \mathbb{R}\}$. This vector space depends on two different scalars, a and b, so it looks like $\mathbb{R}^2$.

Example 1.7

Suppose we look at all infinite sequences $\{a_n\} = (a_0, a_1, a_2, \ldots, a_n, \ldots)$, where $a_n \in \mathbb{R}$. This is a vector space, and looks like $\mathbb{R}^\infty$.

An interesting followup is the set of all infinite sequences with finitely many nonzero objects. This turns out to also be a vector space.

8/27/2021

Lecture 2

Vector Spaces, Subspaces

2.1 Properties of Vector Spaces

Lemma 2.1: Cancellation Law for Vector Addition

If $\vec{x}+\vec{y}=\vec{x}+\vec{z}$, then $\vec{y}=\vec{z}$.

Proof. To cancel the $\vec{x}$'s, we need to add the inverse of $\vec{x}$. Suppose $\vec{x}_1$ is the inverse of $\vec{x}$; then, $\vec{x}+\vec{x}_1=\vec{0}$.

$$\begin{aligned}
\vec{x}+\vec{y}&=\vec{x}+\vec{z} && \\
\vec{x}_1+\vec{x}+\vec{y}&=\vec{x}_1+\vec{x}+\vec{z} && \\
(\vec{x}_1+\vec{x})+\vec{y}&=(\vec{x}_1+\vec{x})+\vec{z} && \text{(assoc.)}\\
\vec{0}+\vec{y}&=\vec{0}+\vec{z} && \text{(add. inv.)}\\
\vec{y}&=\vec{z} && \text{(add. ident.)}
\end{aligned}$$

□

Lemma 2.2: Uniqueness of the Additive Identity

The $\vec{0}$ is unique. That is, there is exactly one $\vec{0}$.

Proof. Let $\vec{0}'$ be another zero vector. All we need to show is that it is equal to the original.

If we add them together, we have
$$\begin{cases}\vec{0}+\vec{0}'=\vec{0}' & (\vec{0}\text{ is zero})\\ \vec{0}+\vec{0}'=\vec{0} & (\vec{0}'\text{ is zero})\end{cases}$$

Since the LHS of both equations are the same, this means that $\vec{0}=\vec{0}'$, and there is exactly one zero vector. □

Lemma 2.3: Uniqueness of the Additive Inverse

The inverses are unique. That is, $(\forall x\in\mathcal{V})(\exists!\vec{y}\in\mathcal{V})(\vec{x}+\vec{y}=\vec{0})$.

Proof. Say $\vec{y}_1$ is some additive inverse of $\vec{x}$. This means that by definition,

$$\vec{x}+\vec{y}=\vec{0}=\vec{x}+\vec{y}_1.$$

Using the cancellation property (Lemma 2.1), we must have that $\vec{y}=\vec{y}_1$. □

Lemma 2.4: Multiplying by 0

$0\cdot\vec{x}=\vec{0}$, where $0\in F$, $\vec{0}\in\mathcal{V}$.

Proof. Suppose $0\cdot\vec{x}=\vec{y}\in\mathcal{V}$. We know $\vec{y}\in\mathcal{V}$ since $\mathcal{V}$ is closed under scalar multiplication.

What happens if we add $\vec{y}$ to itself?

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

We can rewrite this as

$$\begin{aligned} \vec{y} + \vec{y} &= \vec{y} \\ \vec{y} + \vec{y} &= \vec{y} + \vec{0} && \text{(add } \vec{0}\text{)} \\ \vec{y} &= \vec{0} && \text{(cancellation)} \end{aligned}$$

□

Lemma 2.5: Additive Inverse as $-\vec{x}$

The inverse of $\vec{x}$ is $(-1) \cdot \vec{x}$, denoted as $-\vec{x}$.

Proof. Let us check if $(-1)\vec{x}$ is indeed the inverse.

$$\begin{aligned} (-1) \cdot \vec{x} + 1 \cdot \vec{x} &= (-1+1) \cdot \vec{x} && \text{(distr. property)} \\ &= 0 \cdot \vec{x} = \vec{0} && \text{(Lemma 2.4)} \end{aligned}$$

As such, $(-1) \cdot \vec{x} = -\vec{x}$ is the inverse of $\vec{x}$ (and is unique, by Lemma 2.3). □

Lemma 2.6: Multiplying by $\vec{0}$

$a \cdot \vec{0} = \vec{0}$, where $a \in \mathcal{F}$ and $\vec{0} \in \mathcal{V}$.

Proof. Suppose we add the LHS to itself.

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

Using the cancellation property (Lemma 2.1), we have $a \cdot \vec{0} = \vec{0}$. □

Why didn't we include these as axioms? Because it'd be harder to verify whether something is a vector space—we can omit these from the axioms because they can be easily proven from the remaining axioms.

As an aside, recall the notation from last time (Definition 1.1): $\mathcal{F}(S, T)$ is the set of all functions from S to T.

Example 2.7

Is $\mathbb{R}^n = \mathcal{F}(S, T)$ for some S, T?

It turns out, yes! We have that $\mathbb{R}^n$ is just the set of all n-tuples of reals. If we associate each of the indices with one natural number, we have that $\mathbb{R}^n = \mathcal{F}(\{1, 2, \ldots, n\}, \mathbb{R})$.

For example, we can represent $\mathbb{R}^3$ as the set of all functions from $\{1, 2, 3\}$ to $\mathbb{R}$.

2.2 Vector Subspaces

What is the big picture? We have some subobject with the same properties as a big object.

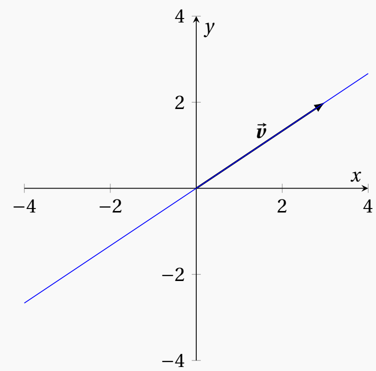
Subsets are an example of this; subfields are another example (Fig. 1). One thing to note is that $\mathbb{N}$ and $\mathbb{Z}$ are not subfields of $\mathbb{Q}$, $\mathbb{R}$ nor $\mathbb{C}$!

Example 2.8: Subspaces of $\mathbb{R}^2$

As a small example, let us focus on the real plane—$\mathbb{R}^2$. What can be a vector subspace of $\mathbb{R}^2$?

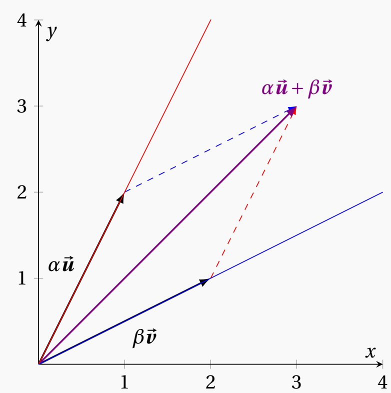
$\mathcal{V} = \mathbb{R}^2$ is a subspace—it's just the entire vector space. But that's kind of boring. What about $W \subset \mathcal{V}$; a proper subspace?

What if we just take a single vector? We'd also need to include all other vectors lying on the same line (as we should be able to multiply by any scalar and stay in the subspace):



We can see that all lines through $\vec{\mathbf{0}}$ are subspaces.

If we add another vector, we just get the subspace $\mathbb{R}^2$ itself—any vector in $\mathbb{R}^2$ can be gotten with a linear combination of $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{v}}$:



The set containing only the zero vector, $\{\vec{\mathbf{0}}\}$, is another subspace—it's the trivial subspace.

An important note to make is that $\varnothing$ is *not* a subspace.

In general,

- We want closure under addition and multiplication
- We want it to contain a zero vector

Definition 2.9: Subspace

Let $\mathcal{V}$ be a vector space over a field F (that is, $\mathcal{V}_F$).

Let $\mathcal{W}$ be a subset of $\mathcal{V}$. Then $\mathcal{W}$ is a vector subspace of $\mathcal{V}$ (i.e. $\mathcal{W} \leq \mathcal{V}$) if $\mathcal{W}$ is itself a vector space over F under $\mathcal{V}$'s operations of vector addition and scalar multiplication.

How do we determine whether a subset $\mathcal{W}$ of $\mathcal{V}$ is a subspace of $\mathcal{V}_F$?

We have two options. The first is with brute force, which is generally very tedious; we need to confirm the definition of a vector space applies. That is, we need to verify all eight vector space axioms, and also that $\mathcal{W}$ is closed.

It turns out that our job can actually be made a little bit easier—we do not need to verify all eight vector space axioms.

Theorem 2.10: Verifying a Vector Subspace

$\mathcal{W}$ is a vector subspace of $\mathcal{V}$ if and only if

- $\vec{\mathbf{0}} \in \mathcal{W}$
- $\mathcal{W}$ is closed under addition
- $\mathcal{W}$ is closed under scalar multiplication

Here, both operations come from $\mathcal{V}$.

Proof. Let us prove the converse (the forward direction follows directly from the definition of subspace).

We need to show that all vector space axioms work in $\mathcal{W}$. It is actually the case that most of them trivially work in $\mathcal{W}$ because they work in $\mathcal{V}$. Specifically, commutativity, associativity, and distributivity all work because those operations have already been verified in $\mathcal{V}$.

The only axioms that do not trivially work are the existence axioms.

Existence of zero vector: Since the zero vector in $\mathcal{V}$ is unique, it must be equal to the zero vector in $\mathcal{W}$ (otherwise, we'd have two zero vectors in $\mathcal{V}$, and this has been disproven in Lemma 2.2).

Existence of inverse: If $\vec{\boldsymbol{w}} \in \mathcal{W}$, then $(-1)\vec{\boldsymbol{w}} \in \mathcal{W}$ because it is closed under scalar multiplication. That means that $-\vec{\boldsymbol{w}}$, the unique inverse, is also in $\mathcal{W}$ (by Lemmas 2.3 and 2.5). □

Corollary 2.11: Nonempty Vector Spaces and the Zero Vector

The first item in Theorem 2.10 can actually be substituted with the requirement that $\mathcal{W}$ is non-empty. That is,

$$\vec{\mathbf{0}}_{\mathcal{V}} \in \mathcal{W} \iff \mathcal{W} \neq \varnothing.$$

Proof. If $\vec{\boldsymbol{w}} \in \mathcal{W}$, then $-\vec{\boldsymbol{w}} \in \mathcal{W}$. This implies that $\vec{\boldsymbol{w}} + (-\vec{\boldsymbol{w}})$ is also in $\mathcal{W}$ because it is closed under addition. That means that $\vec{\mathbf{0}}_{\mathcal{W}} \in \mathcal{W}$. □

It should be noted that every vector space $\mathcal{V}$ has two *trivial* subspaces: the *zero subspace* $\{\vec{\mathbf{0}}\}$ and the *improper subspace* $\mathcal{V}$. However, not every vector space has at least two *distinct* subspaces—the zero vector space is the only counterexample.

8/30/2021

Lecture 3

Examples of Subspaces, Intersections, Sums

Example 3.1

Recall $\mathcal{P}_n(\mathbb{C})$, the vector space of polynomials of degree at most n with complex coefficients.

Let us first define the degree of a constant polynomial c as 0, (if $c \neq 0$), and degree -1 (or $-\infty$) if $c = 0$.

We can define a chain of subspaces: $\mathcal{P}_{-1}(\mathbb{C}) \leq \mathcal{P}_0(\mathbb{C}) \leq \mathcal{P}_1(\mathbb{C}) \leq \cdots \leq \mathcal{P}_n(\mathbb{C}) \leq \cdots \leq \mathcal{P}(\mathbb{C}) = \mathcal{P}_\infty(\mathbb{C})$.

Example 3.2

Recall $\mathcal{F}(\mathbb{R},\mathbb{R})$ as the set of all functions $f : \mathbb{R} \to \mathbb{R}$. Further, recall $\mathscr{D}(\mathbb{R},\mathbb{R})$ as the set of all differentiable functions $f : \mathbb{R} \to \mathbb{R}$. Further, the set of continuous functions $\mathcal{C}(\mathbb{R},\mathbb{R})$.

Thus, we have another chain: $\{\vec{\mathbf{0}}\} \leq \mathcal{P}(\mathbb{R}) \leq \mathscr{D}(\mathbb{R},\mathbb{R}) \leq \mathcal{C}(\mathbb{R},\mathbb{R}) \leq \mathcal{F}(\mathbb{R},\mathbb{R})$

Example 3.3

Let us consider $\mathcal{F}_0(\mathbb{R},\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} \mid f(0) = 0\}$.

We know that the zero vector (zero function) is contained within $\mathcal{F}_0(\mathbb{R},\mathbb{R})$, and we know that it is also closed under scalar multiplication and vector addition, so it is a subspace of $\mathcal{F}(\mathbb{R},\mathbb{R})$.

What about $\mathcal{F}_2(\mathbb{R},\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} \mid f(0) = 2\}$?

The zero vector (zero function) is not in this set, and further, it is not closed under scalar multiplication nor vector addition.

What about $\tilde{\mathcal{F}}_2(\mathbb{R},\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} \mid f(2) = 0\}$?

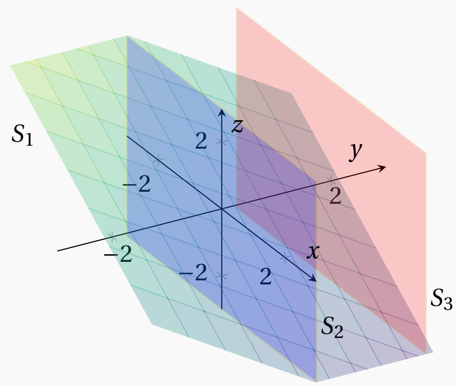
This is also fine, with a similar reasoning as the first—the zero function is in the set, and it is also closed under scalar multiplication and vector addition.

Example 3.4

Remember $\mathbb{R}^3 = \left\{ \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \mid a_i \in \mathbb{R} \right\}$.

Consider $S_1 = \left\{ \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \mid a_i \in \mathbb{R}, a_1 + a_2 + a_3 = 0 \right\}$, and consider $S_2 = \left\{ \begin{bmatrix} a_1 \\ 0 \\ a_3 \end{bmatrix} \mid a_i \in \mathbb{R} \right\}$, and also consider $S_3 = \left\{ \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \mid a_i \in \mathbb{R}, a_2 = 2 \right\}$.

We can find that S_1 and S_2 are subspaces, and S_3 is not. However, all of them geometrically look alike; all three sets are planes in $\mathbb{R}^3$.



This is all fine, but we can do better—we need to find a systematic way to build more subspaces.

The first operation that we can try is to take the intersection:

Theorem 3.5: Intersections of Subspaces

If $\mathcal{W}_1, \mathcal{W}_2 \le \mathcal{V}_F$, then $\mathcal{W}_1 \cap \mathcal{W}_2 \le \mathcal{V}_F$.

Proof. Suppose we are working in $\mathcal{V}_F$, with operations defined in $\mathcal{V}_F$. Since $\vec{\mathbf{0}} \in \mathcal{W}_1, \mathcal{W}_2$, then $\vec{\mathbf{0}} \in \mathcal{W}_1 \cap \mathcal{W}_2$.

Suppose $\vec{\boldsymbol{w}}_1, \vec{\boldsymbol{w}}_2 \in \mathcal{W}_1 \cap \mathcal{W}_2$, and let $c \in F$.

The sum $\vec{\boldsymbol{w}}_1 + \vec{\boldsymbol{w}}_2 \in \mathcal{W}_1$ because both are in $\mathcal{W}_1$. Similarly, it is also in $\mathcal{W}_2$ because of the same reasoning. This means that $\vec{\boldsymbol{w}}_1 + \vec{\boldsymbol{w}}_2 \in \mathcal{W}_1 \cap \mathcal{W}_2$.

Further, the product $c\vec{\boldsymbol{w}}_1 \in \mathcal{W}_1, \mathcal{W}_2$ by a similar reasoning; both are subspaces and closed under scalar multiplication. This means that $c\vec{\boldsymbol{w}}_1 \in \mathcal{W}_1 \cap \mathcal{W}_2$.

All of these put together, we can conclude that $\mathcal{W}_1 \cap \mathcal{W}_2$ is a subspace of $\mathcal{V}_F$. □

Intersection works well, but the union of two subspaces is *not* a subspace. As a simple counterexample, consider two lines in $\mathbb{R}^2$—their union is not a subspace because it is not closed under vector addition.

In fact, the union $\mathcal{W}_1 \cup \mathcal{W}_2$ is almost never a subspace of $\mathcal{V}_F$. It is only a subspace if and only if $\mathcal{W}_1 \subseteq \mathcal{W}_2$ or vice versa.

Example 3.6: Triangular Matrices

Let $\mathcal{U}_{n\times n}(F)$ be the set of all $n \times n$ upper triangular matrices; that is, there are nonzero entries only on or above the diagonal—we have $a_{ij} = 0$ if $i > j$.

Let $\mathcal{L}_{n\times n}(F)$ be the set of all $n \times n$ lower triangular matrices; that is, there are nonzero entries on or below the diagonal—we have $a_{ij} = 0$ if $i < j$.

The intersection $\mathcal{U}_{n\times n}(F) \cap \mathcal{L}_{n\times n}(F)$ is the set of all diagonal matrices—that is, all nonzero entries are along the diagonal. This is a subspace (by the previous theorem).

Definition 3.7: Matrix Transpose

The transpose $\mathbf{A}^T$ of a matrix $\mathbf{A}$ with elements a_{ij} has entries a_{ji}.

Example 3.8: Symmetric Matrices

Let $\mathcal{S}_{n\times n}(F) = \mathcal{S}_n(F)$ be the set of all symmetric matrices—that is, $a_{ij} = a_{ji}$, for all i, j. Alternatively, $\mathbf{A} \in \mathcal{S}_n(F)$ if and only if $\mathbf{A}^T = \mathbf{A}$.

We can find that $\mathcal{S}_n(F)$ is a subspace of all matrices; we have $(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T = \mathbf{A}+\mathbf{B}$, which means that the sum also symmetric. Further, $(c\mathbf{A})^T = c\mathbf{A}^T = c\mathbf{A}$, which means that it is also symmetric.

Example 3.9: Skew-symmetric Matrices

Let $\mathcal{A}_n(F)$ be the set of all skew-symmetric matrices—that is, $a_{ij} = -a_{ji}$, for all i, j. Alternatively, $\mathbf{A} \in \mathcal{A}_n(F)$ if and only if $\mathbf{A}^T = -\mathbf{A}$.

Consider the diagonal—we must have that $a_{i,i} = -a_{i,i}$. In our major fields of $\mathbb{R}$, $\mathbb{C}$, etc., this means that the diagonal must be 0. However, in a field like $\mathbb{Z}_2$, we have $1+1=0$, so the diagonal may not all be zeroes!

In HW, we will show that $\mathcal{A}_n(F) \le \mathcal{M}_{n\times n}(F)$.

Definition 3.10: Sum of Subspaces

We define the sum of two subspaces $\mathcal{W}_1$ and $\mathcal{W}_2$ of $\mathcal{V}$ as

$$\mathcal{W}_1 + \mathcal{W}_2 = \{\vec{\boldsymbol{w}}_1 + \vec{\boldsymbol{w}}_2 \mid \vec{\boldsymbol{w}}_1 \in \mathcal{W}_1, \vec{\boldsymbol{w}}_2 \in \mathcal{W}_2\}.$$

Example 3.11: Sum of Subspaces

Suppose we have $\mathcal{W}_1, \mathcal{W}_2 \le \mathcal{V}$. We want a (minimal) subspace that contains both $\mathcal{W}_1$ and $\mathcal{W}_2$.

We can find that $\mathcal{W}_1 + \mathcal{W}_2 \le \mathcal{V}$, and is in fact the minimal subspace that contains both $\mathcal{W}_1$ and $\mathcal{W}_2$.

The zero element is in $\mathcal{W}_1 + \mathcal{W}_2$, because $\vec{\mathbf{0}}$ is in both $\mathcal{W}_1$ and $\mathcal{W}_2$. We can further show that this set is closed under vector addition and scalar multiplication.

Definition 3.12: Direct Sum

The direct sum $\mathcal{W}_1 \oplus \mathcal{W}_2 = \mathcal{W}_1 + \mathcal{W}_2$ when $\mathcal{W}_1 \cap \mathcal{W}_2 = \{\vec{\mathbf{0}}_{\mathcal{V}}\}$.

Theorem 3.13: Zero Intersections of Subspaces

$\mathcal{W}_1 \cap \mathcal{W}_2 = \{\vec{\mathbf{0}}\}$ if and only if $\forall \vec{\boldsymbol{v}} \in \mathcal{W}_1 + \mathcal{W}_2$ can be uniquely written as $\vec{\boldsymbol{v}} = \vec{\boldsymbol{w}}_1 + \vec{\boldsymbol{w}}_2$ for some $\vec{\boldsymbol{w}}_i \in \mathcal{W}_i$.

9/1/2021

Lecture 4

Linear Combinations, Systems of Equations

Theorem 4.1: Direct Sum of Skew-Symmetric and Symmetric Matrices

$\mathcal{M}_{n\times n}(F) = \mathcal{S}_n(F) \oplus \mathcal{A}_n(F)$, where $F \neq \mathbb{Z}_2$.

Proof. First, we have that $\mathcal{S}_n \cap \mathcal{A}_n = \{\vec{\mathbf{0}}\}$. In order for $\mathbf{A}$ to be both symmetric and skew-symmetric, we must have $-\mathbf{A} = \mathbf{A}^T = \mathbf{A}$. As such, we must have $-\mathbf{A} = \mathbf{A}$ and $\mathbf{A} = \mathbf{0}$.

Next, we need to show that $\mathcal{S}_n + \mathcal{A}_n = \mathcal{M}_{n\times n}$. Suppose $\mathbf{M}_1 \in \mathcal{S}_n$ and $\mathbf{M}_2 \in \mathcal{A}_n$. We then have

$$\begin{aligned}\mathbf{M} &= \mathbf{M}_1 + \mathbf{M}_2 \\ \mathbf{M}^T &= \mathbf{M}_1^T + \mathbf{M}_2^T = \mathbf{M}_1 - \mathbf{M}_2\end{aligned}$$

Solving this system, we have

$$\begin{aligned}\mathbf{M}_1 &= \frac{\mathbf{M} + \mathbf{M}^T}{2} \\ \mathbf{M}_2 &= \frac{\mathbf{M} - \mathbf{M}^T}{2}\end{aligned}$$

To check whether $\mathbf{M}_1$ is symmetric, we have

$$\left(\frac{\mathbf{M} + \mathbf{M}^T}{2}\right)^T = \frac{\mathbf{M}^T + \mathbf{M}}{2}.$$

Similarly,

$$\left(\frac{\mathbf{M} - \mathbf{M}^T}{2}\right)^T = \frac{\mathbf{M}^T - \mathbf{M}}{2},$$

which is the additive inverse of what we started with.

Lastly, we can verify that the sum of $\mathbf{M}_1$ and $\mathbf{M}_2$ is in fact equal to $\mathbf{M}$ (it is). □

4.1 Linear Combinations

Suppose we start with a vector space $\mathcal{V}_F$ and $\vec{v}_1, \vec{v}_2 \in \mathcal{V}$. What kinds of vectors can we get? All vectors of the form $c_1\vec{v}_1 + c_2\vec{v}_2$, where $c_i \in F$.

We can generalize:

Definition 4.2: Linear Combination

If we have $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n \in \mathcal{V}$, then $a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n \in \mathcal{V}$ is a linear combination of $\vec{v}_i$'s with coefficients $a_i \in F$.

What does this mean?

Lemma 4.3: Closure Under Linear Combinations

Every vector space $\mathcal{V}_F$ is closed under linear combinations of its elements.

Further, closure under linear combinations $\iff$ closure under vector addition and scalar multiplication.

Proof. The converse is easily shown (as linear combinations are just vector addition and scalar multiplication), so let us prove the forward direction. Suppose we have $\vec{v}_1, \vec{v}_2 \in \mathcal{V}$, with linear combination $c_1\vec{v}_1 + c_2\vec{v}_2$.

Suppose $c_1 = c_2 = 1$. This gives us $\vec{v}_1 + \vec{v}_2 \in \mathcal{V}$, which is the same as closure under vector addition.

Suppose $c_2 = 0$. This gives us $c_1\vec{v}_2 \in \mathcal{V}$, which is the same as closure under scalar multiplication. □

How can we write down the linear combinations of vectors in an efficient way?

Definition 4.4: Elementary Matrices

An elementary matrix $\mathbf{M}_{kl} = (a_{ij})_{m\times n}$ has elements

$$a_{ij} = \begin{cases} 1 & (i,j) = (k,l) \\ 0 & \text{otherwise} \end{cases}.$$

This means that we can write any matrix $\mathbf{M}$ with elements b_{ij} as

$$\sum_{i,j} b_{ij} \cdot \mathbf{M}_{ij}.$$

4.2 Span

Definition 4.5: Spanning a Vector Space

Vectors $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} \subset \mathcal{V}_F$ span/generate $\mathcal{V}_F$ if every $\vec{v} \in \mathcal{V}$ can be written as a linear combination of the vectors. That is,

$$\vec{v} = a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n \quad \text{for some} \quad a_i \in F.$$

Definition 4.6: Span

In general, the span of several vectors, $\operatorname{span}\{\vec{v}_1, \ldots, \vec{v}_n\}$, is the set of all linear combinations of these vectors. That is,

$$\operatorname{span}\{\vec{v}_1, \ldots, \vec{v}_n\} = \left\{ \sum_{i=1}^{n} a_i\vec{v}_i \mid a_i \in F \right\}.$$

Lemma 4.7: Span as a Subspace

For $\vec{v}_i \in \mathcal{V}$,

$$\operatorname{span}\{\vec{v}_1, \ldots, \vec{v}_n\} \leq \mathcal{V}_F.$$

Proof. The zero vector is in the span if we take $c_i = 0$.

The sum of two linear combinations can be expressed as

$$a_1\vec{v}_1 + \cdots + a_n\vec{v}_n + b_1\vec{v}_1 + \cdots + b_n\vec{v}_n = (a_1 + b_1)\vec{v}_1 + \cdots + (a_n + b_n)\vec{v}_n.$$

All of these coefficients are in the field, so this sum is also in the span.

Similarly, if we multiply a linear combination by some scalar c and distribute, we have another linear combination (full calculation omitted for brevity) □

What is span$\{\varnothing\}$? By definition, let us define span$\{\varnothing\} := \{\vec{\mathbf{0}}\}$. (This is mostly for convenience; we should be able to define the span of an empty set, as it is still a set.)

Example 4.8

Let us take $\mathcal{P}_3(\mathbb{R})$. We can represent this as

$$\mathcal{P}_3(\mathbb{R}) = \operatorname{span}\{1, x, x^2, x^3\}.$$

This is because any degree at most three polynomial can be written as $a \cdot 1 + b \cdot x + c \cdot x^2 + d \cdot x^3$ for $a, b, c, d \in \mathbb{R}$.

Example 4.9

Suppose we have $\vec{\boldsymbol{v}}_1 = 1 + x$ and $\vec{\boldsymbol{v}}_2 = -5x$. What is $\operatorname{span}\{\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2\}$? Intuitively, we think that it is $\mathcal{P}_1(\mathbb{R})$.

To prove set equality, we can prove that the LHS is a subset of the RHS, and vice versa.

We know that the span is a subset of $\mathcal{P}_1(\mathbb{R})$, because we cannot make any higher degree polynomials from a linear combination of $\vec{\boldsymbol{v}}_1$ and $\vec{\boldsymbol{v}}_2$.

To show that $\mathcal{P}_1(\mathbb{R})$ is a subset of the span, suppose we have $\vec{\boldsymbol{v}} = a + bx \in \mathcal{P}_1(\mathbb{R})$.

Suppose that $a + bx = c_1\vec{\boldsymbol{v}}_1 + c_2\vec{\boldsymbol{v}}_2$ for some $c_i \in F$. We then have

$$a + bx = c_1(1 + x) + c_2(-5x) = c_1 + (c_1 - 5c_2)x.$$

This means that $a = c_1$ and $b = c_1 - 5c_2$ and $c_2 = \frac{a-b}{5}$. This means that any vector $\vec{\boldsymbol{v}} = a + bx$ can be written as a linear combination of $\vec{\boldsymbol{v}}_1$ and $\vec{\boldsymbol{v}}_2$—that is, $\vec{\boldsymbol{v}}$ is also in the span of $\vec{\boldsymbol{v}}_1$ and $\vec{\boldsymbol{v}}_2$.

Theorem 4.10: Direct Sums have Unique Sums

If $\mathcal{V} = \mathcal{W}_1 \oplus \mathcal{W}_2$, then every vector $\vec{\boldsymbol{v}} \in \mathcal{V}$ can be uniquely expressed as a sum of two vectors $\vec{\boldsymbol{w}}_1 + \vec{\boldsymbol{w}}_2$ where $\vec{\boldsymbol{w}}_1 \in \mathcal{W}_1$ and $\vec{\boldsymbol{w}}_2 \in \mathcal{W}_2$.

Proof. Suppose $\mathcal{V} = \mathcal{W}_1 \oplus \mathcal{W}_2$. We can now show that every vector in $\mathcal{V}$ can be expressed uniquely as a sum of vectors from $\mathcal{W}_1$ and $\mathcal{W}_2$.

Suppose that we can write $\vec{\boldsymbol{v}} \in \mathcal{W}_1 \oplus \mathcal{W}_2$ as two sums $\vec{\boldsymbol{v}} = \vec{\boldsymbol{w}}_1 + \vec{\boldsymbol{w}}_2$ and $\vec{\boldsymbol{v}} = \vec{\boldsymbol{u}}_1 + \vec{\boldsymbol{u}}_2$. It suffices to show that these two sums must have the same vectors.

Subtracting the two equations, we have $\vec{\mathbf{0}} = (\vec{\boldsymbol{w}}_1 - \vec{\boldsymbol{u}}_1) + (\vec{\boldsymbol{w}}_2 - \vec{\boldsymbol{u}}_2)$. Rearranging, we have $\vec{\boldsymbol{w}}_1 - \vec{\boldsymbol{u}}_1 = \vec{\boldsymbol{u}}_2 - \vec{\boldsymbol{w}}_2$. Notice that the LHS is in $\mathcal{W}_1$ and the RHS is in $\mathcal{W}_2$. This means that we've just constructed a vector in the intersection of $\mathcal{W}_1$ and $\mathcal{W}_2$.

However, by the definition of a direct sum, we must have that $\mathcal{W}_1 \cap \mathcal{W}_2 = \{\vec{\mathbf{0}}\}$, and as such $\vec{\boldsymbol{w}}_1 - \vec{\boldsymbol{u}}_1 = \vec{\boldsymbol{u}}_2 - \vec{\boldsymbol{w}}_2 = \vec{\mathbf{0}}$. In other words, $\vec{\boldsymbol{w}}_1 = \vec{\boldsymbol{u}}_1$ and $\vec{\boldsymbol{w}}_2 = \vec{\boldsymbol{u}}_2$, so the two sums are actually the same. □

9/3/2021

Lecture 5

Linear Independence and Dependence, Bases

5.1 Linear Dependence and Independence

We've talked before about ways in which we can span a vector space $\mathcal{V}$, but today we'll talk about *efficient* ways to span a vector space. For example, $\{1, x, 2 - x\}$ span $\mathcal{P}_1(F)$, but $\{1, x\}$ already do span the vector space.

Definition 5.1: Linear Dependence

In a vector space $\mathcal{V}_F$, the set of vectors $\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2, \ldots, \vec{\boldsymbol{v}}_n \in \mathcal{V}$ is linearly dependent if some non-trivial linear combination of them is $\vec{\mathbf{0}}$. That is, we have

$$c_1\vec{\boldsymbol{v}}_1 + c_2\vec{\boldsymbol{v}}_2 + \cdots + c_n\vec{\boldsymbol{v}}_n = \vec{\mathbf{0}},$$

where some $c_i \neq 0$. This relation is called a *linear relation.*

Suppose we go back to our previous example.

Example 5.2

We know that $\{1, x, 2-x\}$ are linearly dependent because we know that $2-x = 2\cdot 1 + (-1)\cdot x$. This means that our linear relation is

$$2 - x - 2\cdot 1 + 1\cdot x = 0..$$

This is a non-trivial linear relation, so this set is linearly dependent.

Example 5.3

Suppose we have $\vec{\boldsymbol{v}}_1 = \begin{bmatrix}1\\0\end{bmatrix}$, $\vec{\boldsymbol{v}}_2 = \begin{bmatrix}0\\1\end{bmatrix}$, and $\vec{\boldsymbol{v}}_3 = \begin{bmatrix}2\\2\end{bmatrix}$.

We can immediately tell that $\vec{\boldsymbol{v}}_3 = 2\vec{\boldsymbol{v}}_1 + 2\vec{\boldsymbol{v}}_2$, so our non-trivial linear relation is

$$2\vec{\boldsymbol{v}}_1 + 2\vec{\boldsymbol{v}}_2 - \vec{\boldsymbol{v}}_3 = \vec{\mathbf{0}}.$$

Definition 5.4: Linear Independence

In a vector space $\mathcal{V}_F$, the set of vectors $\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2, \ldots, \vec{\boldsymbol{v}}_n \in \mathcal{V}$ is linearly independent if there is no non-trivial linear relation between them.

That is, the set of vectors is linearly independent if the only solution to

$$c_1\vec{\boldsymbol{v}}_1 + c_2\vec{\boldsymbol{v}}_2 + \cdots + c_n\vec{\boldsymbol{v}}_n = \vec{\mathbf{0}}$$

is the trivial solution $c_1 = c_2 = \cdots = c_n = 0$.

Example 5.5

Suppose we want to show that $\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}2\\2\end{bmatrix}\right\}$ is linearly independent.

Our linear relation is

$$c_1\begin{bmatrix}1\\0\end{bmatrix} + c_2\begin{bmatrix}2\\2\end{bmatrix}.$$

Our system of equations is then

$$\begin{cases} c_1 + 2c_2 = 0 \\ \qquad 2c_2 = 0 \end{cases}.$$

We can immediately tell that c_2 must be 0, and as such c_1 must also be 0.

Recall that in earlier examples we first noticed that one vector could be expressed as a linear combination of the others, and from there created the non-trivial linear relation. We can generalize.

Lemma 5.6: Condition for Linear Dependence

The set $\{\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2, \ldots, \vec{\boldsymbol{v}}_n\}$ is linearly dependent if and only if *some* $\vec{\boldsymbol{v}}_i$ is a linear combination of the others.

Proof. ($\Leftarrow$) WLOG, suppose $\vec{\boldsymbol{v}}_1$ can be expressed as a linear combination of the others. that is,

$$\vec{\boldsymbol{v}}_1 = c_2\vec{\boldsymbol{v}}_2 + c_3\vec{\boldsymbol{v}}_3 + \cdots + c_n\vec{\boldsymbol{v}}_n \implies \vec{\boldsymbol{0}} = -\vec{\boldsymbol{v}}_1 + c_2\vec{\boldsymbol{v}}_2 + c_3\vec{\boldsymbol{v}}_3 + \cdots + c_n\vec{\boldsymbol{v}}_n.$$

This is a non-trivial linear relation, so these vectors are dependent.

($\Rightarrow$) If the set of vectors are linearly dependent, then we have $c_1\vec{\boldsymbol{v}}_1 + c_2\vec{\boldsymbol{v}}_2 + \cdots + c_n\vec{\boldsymbol{v}}_n = \vec{\boldsymbol{0}}$, where WLOG $c_1 \neq 0$. This means that c_1^{-1} exists and we can multiply both sides by c_1^{-1} to get

$$\vec{\boldsymbol{v}}_1 = -\frac{c_2}{c_1}\vec{\boldsymbol{v}}_2 - \frac{c_3}{c_1}\vec{\boldsymbol{v}}_3 - \cdots - \frac{c_n}{c_1}\vec{\boldsymbol{v}}_n.$$

$\square$

Example 5.7

Here are some examples of linearly independent and linearly dependent sets.

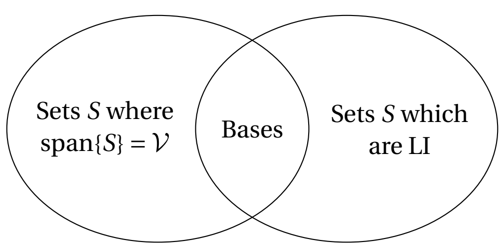
- $\{\vec{\boldsymbol{0}}\}$ is linearly dependent: we have $a \cdot \vec{\boldsymbol{0}} = \vec{\boldsymbol{0}}$ for any scalar $a \neq 0$, which is a non-trivial linear relation.
- The singleton $\{\vec{\boldsymbol{v}} \mid \vec{\boldsymbol{v}} \neq \vec{\boldsymbol{0}}\}$ is linearly independent. Here, suppose there exists a non-trivial linear relation $c \cdot \vec{\boldsymbol{v}} = \vec{\boldsymbol{0}}$, then we have a multiplicative inverse $\vec{\boldsymbol{c}}^{-1}$ and we have $c^{-1}c\vec{\boldsymbol{v}} = \vec{\boldsymbol{0}}$, meaning $\vec{\boldsymbol{v}} = \vec{\boldsymbol{0}}$, which we know is false. By contradiction, c must then equal zero.
- The finite set S where $\vec{\boldsymbol{0}} \in S$ is linearly dependent. Our nontrivial linear relation is then

$$1 \cdot \vec{\boldsymbol{0}} + 0 \cdot \vec{\boldsymbol{v}}_1 + \cdots + 0 \cdot \vec{\boldsymbol{v}}_n.$$

5.2 Bases

Definition 5.8: Basis

A subset $S = \{\vec{\boldsymbol{v}}_1, \ldots, \vec{\boldsymbol{v}}_n\} \subseteq \mathcal{V}_F$ is a *basis* for $\mathcal{V}$ if it both spans $\mathcal{V}$ and is linearly independent.



Example 5.9

The basis for $\mathcal{P}_n(\mathbb{C})$ is the set of $n+1$ elements $\{1, x, x^2, \ldots, x^n\}$.

The basis for $\mathcal{P}(\mathbb{C})$ is the infinite basis $\{1, x, x^2, \ldots, x^n, \ldots\}$.

The basis for $\mathcal{M}_{n\times n}(\mathbb{R})$ is the set of n^2 elementary matrices M_{ij} (that is, the set of matrices with a single 1 at index (i, j)).

We've been counting the number of basis vectors; let us define what this number is.

Definition 5.10: Dimension

The dimension of a vector space $\mathcal{V}$, or $\dim \mathcal{V}$, is the number of vectors in any basis for $\mathcal{V}$.

Example 5.11

Let's look at $\dim\{\vec{\mathbf{0}}\}$. There is no basis, as the set $\{\vec{\mathbf{0}}\}$ is linearly dependent; this means that $\dim\{\vec{\mathbf{0}}\} = 0$.

Example 5.12

What is $\dim \mathcal{S}_{n\times n}(\mathbb{R})$, the set of symmetric matrices?

Any basis matrix with a 1 on (i, j) must also have a 1 at (j, i); that is, $M_{ij} + M_{ji}$ where $i \neq j$ is a basis vector. Further, if $i = j$, then the elementary matrix M_i is also a basis vector.

This set of matrices will always span $\mathcal{M}_{n\times n}(\mathbb{R})$, because any matrix can be uniquely represented as a linear combination of these matrices. These matrices are also linearly independent (proof omitted).

There are a total of $\frac{n(n+1)}{2}$ matrices in this set (this is the number of elements on or above the diagonal).

9/8/2021

Lecture 6

Dimension, Replacement Theorem

Recall the actions we can have on sets:

- S: generating set $\{\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2, \dots, \vec{\boldsymbol{v}}_k\}$ usually has too many elements.

  Dropping some $\vec{\boldsymbol{v}}_j$'s gives us a smaller subset that is linearly independent, forming a basis.
- T: linearly independent set $\{\vec{\boldsymbol{w}}_1, \vec{\boldsymbol{w}}_2, \dots, \vec{\boldsymbol{w}}_m\}$ usually has too few elements.

  Enlarging and adding more vectors gives us a larger subset that is still linearly independent and spans the vector space, forming a basis.

The key idea here is that the number of basis vectors is $\leq k$ and $\geq m$.

Lemma 6.1

If $S = \{\vec{\boldsymbol{v}}_1, \dots, \vec{\boldsymbol{v}}_k\}$ spans $\mathcal{V}$, then some subset S' of S is a basis for $\mathcal{V}$.

Proof. If S is linearly independent, then S is a basis for $\mathcal{V}$.

But if S is not a basis, then S is linearly dependent and WLOG we have $\vec{\boldsymbol{v}}_k = c_1\vec{\boldsymbol{v}}_1 + \cdots + c_{k-1}\vec{\boldsymbol{v}}_{k-1}$.

This means that $\operatorname{span}\{\vec{\boldsymbol{v}}_1, \dots, \vec{\boldsymbol{v}}_{k-1}\} = \operatorname{span}\{\vec{\boldsymbol{v}}_1, \dots, \vec{\boldsymbol{v}}_{k-1}, \vec{\boldsymbol{v}}_k\} = \mathcal{V}$.

If this new smallest subset is not a basis for $\mathcal{V}$, then we can repeat this process. This process can go on for at most k steps—so this will always terminate.

At the end of the process we will have a chain of strict subsets which all span $\mathcal{V}$, with one linearly independent set that spans $\mathcal{V}$; this is the basis for $\mathcal{V}$. $\square$

Lemma 6.2

If S spans $\mathcal{V}$, then $|S| \leq \dim \mathcal{V}$ if $\dim \mathcal{V}$ is well-defined.

Lemma 6.3

If $B = \{\vec{\boldsymbol{v}}_1, \dots, \vec{\boldsymbol{v}}_m\}$ is a basis for $\mathcal{V}$, then any vector $\vec{\boldsymbol{v}} \in \mathcal{V}$ can be written as a unique linear combination of vectors in B.

Theorem 6.4: Replacement Theorem

Suppose we are given $B = \{\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2, \dots, \vec{\boldsymbol{v}}_n\}$ is a basis for $\mathcal{V}$ and $T = \{\vec{\boldsymbol{w}}_1, \dots, \vec{\boldsymbol{w}}_m\}$ is linearly independent in $\mathcal{V}$.

We can extend T to a basis for $\mathcal{V}$ using vectors in B. More precisely, $\exists$ a subset $B_1 \subset B$ such that $T \cup B_1$ is a basis for $\mathcal{V}$ and $|B_1| = n - m$ and $|B_1 \cup T| = n$. In particular, $|T| \leq |B|$.

Proof. We proceed by induction on m (the number of LI vectors), $\forall n \geq 1$.

Base Case ($m = 1$): We have $T = \{\vec{\boldsymbol{w}}_1 \neq \vec{\mathbf{0}}\}$, which is linearly independent, and we have any $n \geq 1$ basis vectors in B.

Since we have a basis, we can express
$$\vec{\boldsymbol{w}}_1 = c_1 \vec{\boldsymbol{v}}_1 + \cdots + c_n \vec{\boldsymbol{v}}_n.$$
Since $\vec{\boldsymbol{w}}_1 \neq \vec{\mathbf{0}}$, we must have at least one $c_i \neq 0$. WLOG, assume $c_n \neq 0$. We then then solve for $\vec{\boldsymbol{v}}_n$, which we now know to be in $\operatorname{span}\left\{\vec{\boldsymbol{w}}_1, \vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2, \dots, \vec{\boldsymbol{v}}_{n-1}\right\}$.

This set is now a basis for $\mathcal{V}$; it spans $\mathcal{V}$ and is also linearly independent (these proofs are for HW).

Inductive Hypothesis: Assume that the claim is true for some $m \geq 1$ (for all $n \geq 1$).

Inductive Step: We now show the claim is true for $m + 1$ (for all $n \geq 1$).

Let $T = \{\vec{\boldsymbol{w}}_1, \dots, \vec{\boldsymbol{w}}_m, \vec{\boldsymbol{w}}_{m+1}\}$ be a linearly independent set. We want to expand this to a basis for $\mathcal{V}$. However, we already know that we can do this for $T' = \{\vec{\boldsymbol{w}}_1, \dots, \vec{\boldsymbol{w}}_m\}$. Further, we know that this smaller set is linearly independent, and by the IH we can expand T' to a basis for $\mathcal{V}$:

$$T'' = \{\vec{\boldsymbol{w}}_1, \dots, \vec{\boldsymbol{w}}_m, \vec{\boldsymbol{v}}_1, \dots, \vec{\boldsymbol{v}}_{n-m}\}.$$

(Notice that we also added $n - m$ basis vectors, which means that we now have a size n basis.)

Since T'' is a basis, we can express $\vec{\boldsymbol{w}}_{m+1}$ as a linear combination of elements in T'':

$$\vec{\boldsymbol{w}}_{m+1} = c_1 \vec{\boldsymbol{w}}_1 + \cdots + c_m \vec{\boldsymbol{w}}_m + b_1 \vec{\boldsymbol{v}}_1 + \cdots + b_{n-m} \vec{\boldsymbol{v}}_1 + \cdots + b_{n-m} \vec{\boldsymbol{v}}_{n-m}.$$

We know that some $b_i \neq 0$ because otherwise $\vec{\boldsymbol{w}}_{m+1}$ would be in the span of the others in T (which we know to be false, as it is linearly independent). This means that WLOG we can assume $b_{n-m} \neq 0$. We can then solve for $\vec{\boldsymbol{v}}_{n-m}$, and replace it with the linear combination involving $\vec{\boldsymbol{w}}_{m+1}$.

This is also a basis for $\mathcal{V}$, as it still spans $\mathcal{V}$ and is also linearly independent (proofs are left for HW).

As such, by the principles of induction, the claim holds for all $n, m \geq 1$. $\square$

Lemma 6.5

If $\mathcal{V}$ has a finite basis, then all of its bases have number of vectors.

Proof. Let S_1 and S_2 be two bases for $\mathcal{V}$. We have

- S_1 is linearly independent, S_2 is a basis for $\mathcal{V}$.

  By the replacement theorem, we can then extend S_1 to become a basis, with the same number of elements as S_2; that is, $|S_1| \le |S_2|$.

- S_2 is linearly independent, S_1 is a basis for $\mathcal{V}$.

  By the replacement theorem, we can then extend S_2 to become a basis, with the same number of elements as S_1; that is, $|S_2| \le |S_1|$.

Putting this together, we know that $|S_1| = |S_2|$. □

Here are some powerful shortcuts to determine whether a set is a basis for $\mathcal{V}$. Suppose $\dim \mathcal{V} = n$.

- If S is linearly independent and $|S| = n$, then S is a basis for $\mathcal{V}$.
- If S spans $\mathcal{V}$ and $|S| = n$, then S is a basis for $\mathcal{V}$.

6.1 Applications to Subspaces

Suppose we have $\mathcal{W} \le \mathcal{V}$ (that is, $\mathcal{W}$ is a subspace of $\mathcal{V}$), where $\dim \mathcal{V} < \infty$. We then have

- $\dim \mathcal{W} \le \dim \mathcal{V}$.

  We can extend any basis for $\mathcal{W}$ to create a basis for $\mathcal{V}$ because it is already linearly independent in $\mathcal{V}$ (but might not span $\mathcal{V}$).

- $\mathcal{W} = \mathcal{V}$ if and only if $\dim \mathcal{W} = \dim \mathcal{V}$.

  Similarly, if we start with a basis for $\mathcal{W}$, it must also be a basis for $\mathcal{V}$ because it is linearly independent and is of the right size.

- Suppose $\mathcal{W}_1, \mathcal{W}_2 \le \mathcal{V}$ (i.e. $\mathcal{W}_1, \mathcal{W}_2$ are subspaces of $\mathcal{V}$).

  We have that $\mathcal{W}_1 + \mathcal{W}_2, \mathcal{W} \cap \mathcal{W}_2 \le \mathcal{V}$.

  This means that $\dim(\mathcal{W}_1 \cap \mathcal{W}_2) \le \dim \mathcal{W}_1$ and $\dim \mathcal{W}_2 \le \dim(\mathcal{W}_1 + \mathcal{W}_2)$.

  Further, $\dim(\mathcal{W}_1 + \mathcal{W}_2) = \dim \mathcal{W}_1 + \dim \mathcal{W}_2 - \dim(\mathcal{W}_1 \cap \mathcal{W}_2)$.

  This reminds us of $|S_1 \cup S_2| = |W_1| + |W_2| - |W_1 \cap W_2|$. Our previous formula is almost the same, except we have a sum instead of a union.

Example 6.6

What is $\dim(\mathcal{U}_n + \mathcal{L}_n)$?

We know that $\dim \mathcal{U}_n = \dim \mathcal{L}_n = \frac{n(n+1)}{2}$, and that $\dim \mathcal{U}_n \cap \mathcal{L}_n = n$ because the intersection is the set of diagonal matrices.

We know that any matrix can be written as the sum of an upper and lower triangular matrix, so we have

$$\frac{n(n+1)}{2} + \frac{n(n+1)}{2} - n = n^2.$$

This verifies these formulas.

9/10/2021

Lecture 7

Linear Transformations: Kernels and Images

7.1 Linear Transformations

As background, we're working with a category of vector spaces. Relationships between objects (vector spaces) within the category are called maps, and linear transformations.

One question is: what is a "good" function? A function that preserves/respects vector space operations.

Definition 7.1: Linear Transformation

Let $\mathcal{V}, \mathcal{W}$ be vector spaces over the same field F. A map $T : \mathcal{V} \to \mathcal{W}$ is a linear transformation if:

1. $T(\vec{\boldsymbol{v}}_1 + \vec{\boldsymbol{v}}_2) = T(\vec{\boldsymbol{v}}_1) + T(\vec{\boldsymbol{v}}_2)$, $\forall \vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2 \in \mathcal{V}$
2. $T(c\vec{\boldsymbol{v}}) = cT(\vec{\boldsymbol{v}})$, $\forall \vec{\boldsymbol{v}} \in \mathcal{V}$, $c \in F$

Example 7.2

Let $\mathcal{D} : P_n(\mathbb{R}) \to P_n(\mathbb{R})$ is the differential operator. That is, we have a map from $f(x) \mapsto f'(x)$.

We can check whether it respects addition and scalar multiplication:

1. $\mathcal{D}(f+g) = (f+g)' = f' + g' = \mathcal{D}(f) + \mathcal{D}(g)$
2. $\mathcal{D}(cf) = (cf)' = cf' = c\mathcal{D}(F)$

Since both operations are preserved, this is a linear transformation.

7.2 Properties of Linear Transformations

Suppose $T : \mathcal{V} \to \mathcal{W}$ is a linear transformation.

1. T preserves $\vec{\mathbf{0}}$. That is, $T(\vec{\mathbf{0}}_{\mathcal{V}}) = T(\vec{\mathbf{0}}_{\mathcal{W}})$.

   We can see that this is true because

   $$T(\vec{\mathbf{0}}_{\mathcal{V}}) = T(\vec{\mathbf{0}}_{\mathcal{V}} + \vec{\mathbf{0}}_{\mathcal{V}}) = T(\vec{\mathbf{0}}_{\mathcal{V}}) + T(\vec{\mathbf{0}}_{\mathcal{V}}).$$

   Canceling out, we have that $\vec{\mathbf{0}}_{\mathcal{W}} = T(\vec{\mathbf{0}}_{\mathcal{V}})$

2. T preserves linear combinations. That is, $T(c_1\vec{\boldsymbol{v}}_1 + \cdots + c_n\vec{\boldsymbol{v}}_n) = c_1T(\vec{\boldsymbol{v}}_1) + \cdots + c_nT(\vec{\boldsymbol{v}}_n)$.

   We can prove this by induction, and is left as an exercise on the homework.

One insight is that instead of checking both preservation of vector addition and scalar multiplication, we can just check preservation of linear combinations. Both definitions will follow.

7.3 Distinguished Subspaces

Definition 7.3: Kernel

The nullspace (Kernel) of $T : \mathcal{V} \to \mathcal{W}$ is the set of all vectors $\vec{\boldsymbol{v}} \in \mathcal{V}$ that are sent to $\vec{\mathbf{0}}_{\mathcal{W}}$ by T.

$$\operatorname{Ker} T = \mathrm{N}(T) = \{\vec{\boldsymbol{v}} \in \mathcal{V} \mid T(\vec{\boldsymbol{v}}) = \vec{\mathbf{0}}_{\mathcal{W}}\}.$$

Lemma 7.4

$\operatorname{Ker} T \leqq \mathcal{V}$.

Proof. We know $\vec{\mathbf{0}}_{\mathcal{V}} \in \operatorname{Ker} T$ because $T(\vec{\mathbf{0}}_{\mathcal{V}}) = \vec{\mathbf{0}}_{\mathcal{W}}$.

For $\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2 \in \operatorname{Ker} T$ and $c_1, c_2 \in F$, we also have

$$T(c_1 \vec{\boldsymbol{v}}_1 + c_2 \vec{\boldsymbol{v}}_2) = c_1 T(\vec{\boldsymbol{v}}_1) + c_2 T(\vec{\boldsymbol{v}}_2) = \vec{\mathbf{0}}_{\mathcal{W}}$$

As such, $\operatorname{Ker} T$ is a subspace of $\mathcal{V}$. □

Definition 7.5: Image

The range (Image) of $T : \mathcal{V} \to \mathcal{W}$ is the set

$$\mathrm{R}(T) = \operatorname{Im} T = \{\vec{\boldsymbol{w}} \in \mathcal{W} \mid \vec{\boldsymbol{w}} = T(\vec{\boldsymbol{v}}) \text{ for some } \vec{\boldsymbol{v}} \in \mathcal{V}\}.$$

Example 7.6

$\operatorname{Im}(\mathcal{D} : P_n(\mathbb{R}) \to P_n(\mathbb{R})) = P_{n-1}(\mathbb{R})$

$\operatorname{Ker}(\mathcal{D} : P_n(\mathbb{R}) \to P_n(\mathbb{R})) = P_0(\mathbb{R})$

Lemma 7.7

$\operatorname{Im} T \leqq \mathcal{W}$.

Proof. $\vec{\mathbf{0}}_{\mathcal{W}} \in \operatorname{Im} T$ as $T(\vec{\mathbf{0}}_{\mathcal{V}}) = \vec{\mathbf{0}}_{\mathcal{W}}$.

For $\vec{\boldsymbol{w}}_1, \vec{\boldsymbol{w}}_2 \in \operatorname{Im} T$ and $c_1, c_2 \in F$, we have $c_1 \vec{\boldsymbol{w}}_1 + c_2 \vec{\boldsymbol{w}}_2 \in \operatorname{Im} T$.

This is because we must have some $T(\vec{\boldsymbol{v}}_1) = \vec{\boldsymbol{w}}_1$ and $T(\vec{\boldsymbol{v}}_2) = \vec{\boldsymbol{w}}_2$, and $T(c_1 \vec{\boldsymbol{v}}_1 + c_2 \vec{\boldsymbol{v}}_2) = c_1 T(\vec{\boldsymbol{v}}_1) + c_2 T(\vec{\boldsymbol{v}}_2) = c_1 \vec{\boldsymbol{w}}_1 + c_2 \vec{\boldsymbol{w}}_2$.

As such, $\operatorname{Im} T$ is a subspace of $\mathcal{W}$. □

We have $\dim \mathrm{N}(T) = \dim \operatorname{Ker} T = \operatorname{null}(T)$ and $\dim \mathrm{R}(T) = \dim \operatorname{Im} T = \operatorname{rank}(T)$.

Theorem 7.8: Dimension Theorem

For $T : \mathcal{V} \to \mathcal{W}$, $\dim(\operatorname{Ker} T) + \dim(\operatorname{Im} T) = \dim \mathcal{V}$.

Proof. Suppose we start from $\operatorname{Ker} T$. Let $\{\vec{\boldsymbol{v}}_1, \dots, \vec{\boldsymbol{v}}_k\}$ be a basis for $\operatorname{Ker} T$.

By the Replacement Theorem, we can extend this basis to a basis for $\mathcal{V}$; that is, we add $\{\vec{\boldsymbol{v}}_{k+1}, \dots \vec{\boldsymbol{v}}_n\}$ to our existing basis.

If we apply the linear transformation T, the vectors $\{\vec{\boldsymbol{v}}_1, \dots, \vec{\boldsymbol{v}}_k\}$ will all map to $\vec{\mathbf{0}}$, and as such we'll end up with only $\{\vec{\boldsymbol{v}}_{k+1}, \dots, \vec{\boldsymbol{v}}_n\}$.

We know that since $\{\vec{\boldsymbol{v}}_1, \dots, \vec{\boldsymbol{v}}_n\}$ are a basis for $\mathcal{V}$, then $T(\vec{\boldsymbol{v}}_1), \dots, T(\vec{\boldsymbol{v}}_n)$ are a basis for $\operatorname{Im} T$. Since we only have $n - k$ nonzero vectors after the transformation, then $\dim \operatorname{Im} T = n - k$.

We can now show that $\{T(\vec{\boldsymbol{v}}_{k+1}, \dots, T(\vec{\boldsymbol{v}}_n)\}$ are linearly independent. With a linear relation, we have

$$a_{k+1} T(\vec{\boldsymbol{v}}_{k+1}) + \cdots + a_n T(\vec{\boldsymbol{v}}_n) = \vec{\mathbf{0}}$$

$$T(a_{k+1}\vec{\boldsymbol{v}}_{k+1} + \cdots + a_n\vec{\boldsymbol{v}}_n) = \vec{\mathbf{0}}$$

As such, the input $a_{k+1}\vec{\boldsymbol{v}}_{k+1} + \cdots + a_n\vec{\boldsymbol{v}}_n \in \operatorname{Ker} T$, and as such must be equal to some linear combination of the basis vectors of the kernel.

However, this can't be possible, as we know that $\{\vec{\boldsymbol{v}}_1, \dots, \vec{\boldsymbol{v}}_n\}$ are linearly independent (they're a basis for $\mathcal{V}$), and as such all coefficients are 0. This means that $\{T(\vec{\boldsymbol{v}}_{k+1}), \dots, T(\vec{\boldsymbol{v}}_n)\}$ are linearly independent. □

Example 7.9

We have $\operatorname{null}(\mathcal{D}) = \dim(\operatorname{Ker} T) = \dim(P_0(\mathbb{R})) = 1$ and $\operatorname{rank}(\mathcal{D}) = \dim(\operatorname{Im} T) = \dim(P_{n-1}(\mathbb{R})) = n-1$.

Adding them up, we have $\dim P_n(\mathbb{R}) = 1 + n - 1 = n$.

Lemma 7.10

For $T : \mathcal{V} \to \mathcal{W}$, $\operatorname{Im} T = \operatorname{span}\{T(\vec{\boldsymbol{v}}_1), \dots, T(\vec{\boldsymbol{v}}_n)\}$ where $\{\vec{\boldsymbol{v}}_1, \dots, \vec{\boldsymbol{v}}_n\}$ is a basis for $\mathcal{V}$.

Proof. We have $\vec{\boldsymbol{w}} \in \operatorname{Im} T$ iff $T(\vec{\boldsymbol{v}}) = \vec{\boldsymbol{w}}$ for some $\vec{\boldsymbol{v}} \in \mathcal{V}$.

We can then rewrite $\vec{\boldsymbol{v}} = c_1\vec{\boldsymbol{v}}_1 + \cdots + c_n\vec{\boldsymbol{v}}_n$ for some $c_i \in F$ and basis vectors $\vec{\boldsymbol{v}}_i$.

Using linearity, we have $T(\vec{\boldsymbol{v}}) = c_1 T(\vec{\boldsymbol{v}}_1) + \cdots + c_n T(\vec{\boldsymbol{v}}_n)$, which describes any $\vec{\boldsymbol{w}} = T(\vec{\boldsymbol{v}})$.

This means that $\mathcal{W} = \operatorname{span}\{T(\vec{\boldsymbol{v}}_1), \dots, T(\vec{\boldsymbol{v}}_n)\}$. □

A warning: if we start with a basis, T may not preserve linear independence—that is, $\{T(\vec{\boldsymbol{v}}_1, \dots, T(\vec{\boldsymbol{v}}_n)\}$ may not be linearly independent.

9/13/2021

Lecture 8

Properties and Matrix of a Linear Transformation

8.1 Properties of Functions

Definition 8.1: One-to-one (injective)

A function is *injective* if $f(x) = f(y)$ then $x = y$.

Definition 8.2: Onto (surjective)

A function is *surjective* if $\operatorname{Im} f = B$, for $f : A \to B$.

Theorem 8.3: One-to-one Via Kernel

T is one-to-one iff $\operatorname{Ker} T = \{\vec{\mathbf{0}}\}$.

That is, T must have the smallest possible kernel.

Proof. ($\Longrightarrow$) If T is one-to-one, then $T(\vec{\mathbf{0}}_{\mathcal{V}}) = \vec{\mathbf{0}}_{\mathcal{W}}$, and $T(\vec{\boldsymbol{x}}) = \vec{\mathbf{0}}_{\mathcal{W}}$ means that $\vec{\boldsymbol{x}} = \vec{\mathbf{0}}_{\mathcal{V}}$.

This means that $\operatorname{Ker} T = \{\vec{\mathbf{0}}_{\mathcal{V}}\}$.

$(\Longleftarrow)$ If $\operatorname{Ker} T = \{\vec{\mathbf{0}}_{\mathcal{V}}\}$, then $T(\vec{\boldsymbol{x}}) = T(\vec{\boldsymbol{y}})$ means that $T(\vec{\boldsymbol{x}}) - T(\vec{\boldsymbol{y}}) = \vec{\mathbf{0}}$.

Hence, since T is linear, $T(\vec{\boldsymbol{x}} - \vec{\boldsymbol{y}}) = \vec{\mathbf{0}}$. Recalling the fact that the kernel is $\{\vec{\mathbf{0}}\}$, then because $\vec{\boldsymbol{x}} - \vec{\boldsymbol{y}} = \vec{\mathbf{0}}$, then $\vec{\boldsymbol{x}} - \vec{\boldsymbol{y}} = \vec{\mathbf{0}}_{\mathcal{V}}$ and $\vec{\boldsymbol{x}} = \vec{\boldsymbol{y}}$

Thus, by the definition of one-to-one, T is one-to-one. □

Theorem 8.4: Onto via Image

T is onto iff $\operatorname{Im} T = \mathcal{W}$ for $T : \mathcal{V} \to \mathcal{W}$.

That is, T must have largest possible image.

Here are some (potentially) burning questions:

1. What is the minimum information that determines uniquely a linear transformation $T : \mathcal{V} \to \mathcal{V}$?

   If we know where T sends a basis of $\mathcal{V}$, then we know T.

   Suppose we take a basis $\beta = \{\vec{\boldsymbol{v}}_1, \dots, \vec{\boldsymbol{v}}_n\}$, which gets sent to $\{T(\vec{\boldsymbol{v}}_1), \dots, T(\vec{\boldsymbol{v}}_n)\} \in \mathcal{W}$.

   We can *linearly extend* T: That is, for $\vec{\boldsymbol{v}} = c_1\vec{\boldsymbol{v}}_1 + \cdots + c_n\vec{\boldsymbol{v}}_n \in \mathcal{V}$, then $T(\vec{\boldsymbol{v}} = c_1 T(\vec{\boldsymbol{v}}_1) + \cdots + c_n T(\vec{\boldsymbol{v}}_n)$.

2. How do we prove whether a function $T : \mathcal{V} \to \mathcal{W}$ is linear or not?

   To show that something is not linear, we search for any (simple) counterexample that violates the definition or any properties of T.

3. What happens with linearly independent, spanning, bases, linearly dependent sets under T?

Theorem 8.5: Pre-image Preserves Linear Independence

A pre-image of a linearly independent set in $\operatorname{Im} T$ is linearly independent in $\mathcal{V}$.

Proof. If we have a linearly independent set $\{\vec{\boldsymbol{w}}_i\} \in \operatorname{Im} T$, we have to look at the coefficients of the linear relation $c_1\vec{\boldsymbol{v}}_1 + \cdots + c_k\vec{\boldsymbol{v}}_k = \vec{\mathbf{0}}$.

If we apply T, we have

$$\begin{aligned} T(c_1\vec{\boldsymbol{v}}_1 + \cdots + c_k\vec{\boldsymbol{v}}_k) &= T(\vec{\mathbf{0}}) \\ c_1 T(\vec{\boldsymbol{v}}_1) + \cdots + c_k T(\vec{\boldsymbol{v}}_k) &= \vec{\mathbf{0}} \\ c_1\vec{\boldsymbol{w}}_1 + \cdots + c_k\vec{\boldsymbol{w}}_k &= \vec{\mathbf{0}} \end{aligned}$$

Since $\vec{\boldsymbol{w}}_i$'s were linearly independent, then all the coefficients are all zero, and $\{\vec{\boldsymbol{v}}_i\}$ are all linearly independent. □

However, note that T may destroy linear independence—take $T(\vec{\boldsymbol{x}}) = \vec{\mathbf{0}}$.

However, what must our linear transformation be in order to preserve linear independence?

Theorem 8.6: One-to-one Preserves Linear Independence

If T is one-to-one, then T preserves linear independence.

Proof. Suppose we have $\{\vec{\boldsymbol{v}}_i\}$ in $\mathcal{V}$ and $\{\vec{\boldsymbol{w}}_i\}$ in $\mathcal{W}$ for $T : \mathcal{V} \to \mathcal{W}$.

$$\sum_i c_i \vec{\boldsymbol{w}}_i = \vec{\mathbf{0}}$$
$$\sum_i c_i T(\vec{\boldsymbol{v}}_i) = \vec{\mathbf{0}}$$
$$T\left(\sum_i c_i \vec{\boldsymbol{v}}_I\right) = \vec{\mathbf{0}}$$
$$\sum_i c_i \vec{\boldsymbol{v}}_i = \vec{\mathbf{0}}$$

Since we know that $\vec{\boldsymbol{v}}_i$ were linearly independent, then we must have that $c_i = 0$ and $\{\vec{\boldsymbol{w}}_i\}$ is linearly independent. $\square$

LI	$\to$	?
LI	$\leftarrow$	LI
Basis	$\to$	spans
LI	$\leftarrow$	Basis
spans	$\to$	spans
?	$\leftarrow$	spans
LD	$\to$	LD
?	$\leftarrow$	LD

8.2 Coordinates

Suppose we have $\mathcal{V}_{\mathbb{R}}$ where $\dim \mathcal{V} = n$, with $\{\vec{\boldsymbol{v}}_1, \ldots, \vec{\boldsymbol{v}}_n\}$ as a basis for $\mathcal{V}$.

IF we have a basis of n vectors, we can think of $\mathcal{V}$ as "looking like" $\mathbb{R}^n$. That is, there is a one-to-one correspondence (that is, one-to-one and onto, i.e. a bijection) between $\mathcal{V}$ and $\mathbb{R}^n$. Formally,

$$\mathcal{V} \ni \vec{\boldsymbol{v}} = a_1\vec{\boldsymbol{v}}_1 + \cdots + a_n\vec{\boldsymbol{v}}_n \mapsto \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \in \mathbb{R}^n.$$

This is in fact a function from $\mathcal{V}$ to $\mathbb{R}^n$. However, note that this works as long as we first *fix* and *order* the basis. If we reorder the basis, we get a different function! Further, we can go backwards through this function because it is bijective.

Definition 8.7: Coordinates

The *coordinates* of $\vec{\boldsymbol{v}}$ with respect to a basis $\alpha = \{\vec{\boldsymbol{v}}_1, \ldots, \vec{\boldsymbol{v}}_n\}$ are the unique scalars $c_1, c_2, \ldots, c_n$ such that $\vec{\boldsymbol{v}} = c_1\vec{\boldsymbol{v}}_1 + c_2\vec{\boldsymbol{v}}_2 + \cdots + c_n\vec{\boldsymbol{v}}_n$.

The question is whether the coordinate map $T : \mathcal{V} \to \mathbb{R}^n$ linear? We can easily show that $T(\vec{\boldsymbol{v}} + \vec{\boldsymbol{w}}) = T(\vec{\boldsymbol{v}}) + T(\vec{\boldsymbol{w}})$ and that $T(c\vec{\boldsymbol{v}}) = cT(\vec{\boldsymbol{v}})$. (Calculation omitted for brevity).

Later on, we will show that T is an *isomorphism* (linear, one-to-one, and onto).

8.3 Matrix of a Linear Transformation

Suppose we fix two bases $\beta = \{\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2, \ldots, \vec{\boldsymbol{v}}_n\} \in \mathcal{V}$ and $\gamma = \{\vec{\boldsymbol{w}}_1, \vec{\boldsymbol{w}}_1, \ldots, \vec{\boldsymbol{w}}_m\} \in \mathcal{W}$ for $T : \mathcal{V} \to \mathcal{W}$.

Recall that we only need to know $T(\beta)$ to know T. We know that $T(\vec{\boldsymbol{v}}_1), \ldots, T(\vec{\boldsymbol{v}}_n)$ maps to the basis γ in $\mathcal{W}$.

Definition 8.8: Coordinate Vector

The *coordinate vector* of $\vec{v} \in \mathcal{V}$ with respect to the basis α of $\mathcal{V}$ is

$$[\vec{v}]_\alpha = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}_\alpha .$$

Lemma 8.9

If $\dim \mathcal{V} = n$ and $\beta = \{\vec{v}_1, \ldots, \vec{v}_n\}$ is an ordered basis for $\mathcal{V}$, the function $T : \mathcal{V} \to \mathbb{R}^n$

$$\vec{v} \mapsto \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = [\vec{v}]_\beta$$

is a linear transformation

What do we know about this linear transformation $T : \mathcal{V}^\beta \to \mathcal{W}^\gamma$? Suppose we fix the bases $\beta = \{\vec{v}_i\}$, $\gamma = \{\vec{w}_i\}$. $[T(\vec{v}_1)]_\gamma, \ldots, [T(\vec{v}_n)]_\gamma$ contain all our info needed.

Suppose we wend our original basis vectors into $\mathcal{W}$.

$$\begin{aligned} T(\vec{v}_1) &= a_{11}\vec{w}_1 + a_{21}\vec{w}_2 + \cdots + a_{m1}\vec{w}_m && \text{1st column} \\ T(\vec{v}_2) &= a_{12}\vec{w}_1 + a_{22}\vec{w}_2 + \cdots + a_{m2}\vec{w}_m && \text{2nd column} \\ &\vdots \\ T(\vec{v}_n) &= a_{1n}\vec{w}_1 + a_{2n}\vec{w}_2 + \cdots + a_{mn}\vec{w}_m && n\text{th column} \end{aligned}$$

This means that we have

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m_1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

That is, the columns are the coordinate vectors of the images of the basis vectors $\vec{v}_1, \ldots, \vec{v}_n$ of $\mathcal{V}$.

Definition 8.10: Matrix of a Transformation

The matrix of $T : \mathcal{V} \to \mathcal{W}$ with respect to the basis β for $\mathcal{V}$ and the basis γ for $\mathcal{W}$ is

$$[T]_\beta^\gamma = \begin{bmatrix} | & | & & | \\ [T(\vec{v}_1)]_\gamma & [T(\vec{v}_2)]_\gamma & \cdots & [T(\vec{v}_n)]_\gamma \\ | & | & & | \end{bmatrix},$$

where the columns are the images of the basis β of $\mathcal{V}$ written in the basis γ of $\mathcal{W}$.

Note that $[T]_\beta^\gamma$ is of size $m \times n = \dim \mathcal{W} \times \dim \mathcal{V}$.

Example 8.11

Suppose we have $T: \mathbb{R}^3 \to \mathbb{R}^2$ such that $(a_1, a_2, a_3) \mapsto (a_2 + a_3, -a_1 + 2a_2)$.

We have that $3 = \dim \mathcal{V}$, the number of columns, and $2 = \dim \mathcal{W}$, the number of rows.

Suppose we have the bases

$$\beta = \{\vec{\boldsymbol{e}}_1, \vec{\boldsymbol{e}}_2, \vec{\boldsymbol{e}}_3\} \qquad \gamma = \{\vec{\boldsymbol{e}}_1, \vec{\boldsymbol{e}}_2\}$$

We have the following:

$$\begin{aligned} T(\vec{\boldsymbol{e}}_1) &= T(1,0,0) = (0,-1) = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \\ T(\vec{\boldsymbol{e}}_2) &= T(0,1,0) = (1,2) = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \\ T(\vec{\boldsymbol{e}}_3) &= T(0,0,1) = (1,0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \end{aligned}$$

This means that

$$[T]_\beta^\gamma = \begin{bmatrix} 0 & 1 & 1 \\ -1 & 2 & 0 \end{bmatrix}$$

is the 2×3 matrix for T.

9/15/2021

Lecture 9

Matrices of Linear Combinations, linear Transformations, Matrix Multiplication

Theorem 9.1: Image of Linear Transformation

The image of $\vec{\boldsymbol{v}} = [\vec{\boldsymbol{v}}]_\beta$ under $T: \mathcal{V}_\beta^n \to \mathcal{W}_\gamma^m$ is the product of the matrix of $[T]_\beta^\gamma$ with the coordinate vector:

$$[T(\vec{\boldsymbol{v}})]_\gamma = [T]_\beta^\gamma \cdot [\vec{\boldsymbol{v}}]_\beta.$$

How do we multiply matrices?

Definition 9.2: Matrix Multiplication

If we have $\mathbf{A}$ of size $m \times n$ and $\mathbf{B}$ of size $n \times k$, then the product $\mathbf{A} \cdot \mathbf{B} = \mathbf{C}$ of size $m \times k$ where each element

$$c_{ij} = (i\text{th row of } \mathbf{A}) \cdot (j\text{th column of } \mathbf{B}) = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}.$$

Suppose we have some linear transformation $T: \mathcal{V}_\beta^n \to \mathcal{W}_\gamma^m$. We can describe this transformation using a matrix $\mathbf{M} = [T]_\beta^\gamma \in \mathcal{M}_{m \times n}(F)$.

This means that we have a mapping between linear transformations T and $m \times n$ matrices $\mathbf{M}$.

The same idea holds in reverse; given a matrix $\mathbf{M} \in \mathcal{M}_{m \times n}(F)$, there is a unique linear transformation $T: \mathcal{V} \to \mathcal{W}$ whose matrix with respect to β and γ is $\mathbf{M}$.

Example 9.3

Given $\mathbf{M} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \in \mathcal{M}_{2\times 3}(\mathbb{R})$, what is $T : \mathbb{R}^3 \to \mathbb{R}^2$ with matrix $\mathbf{M}$?

We know that

$$T(\vec{\boldsymbol{e}}_1) = \begin{bmatrix} 1 \\ 4 \end{bmatrix} \qquad T(\vec{\boldsymbol{e}}_2) = \begin{bmatrix} 2 \\ 5 \end{bmatrix} \qquad T(\vec{\boldsymbol{e}}_3) = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

In other words,

$$T(a_1, a_2, a_3) = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} a_1 + 2a_2 + 3a_3 \\ 4a_1 + 6a_2 + 6a_3 \end{bmatrix}.$$

If we look at $\mathcal{L}(\mathcal{V}, \mathcal{W})$, we can define this as the set of all linear transformations $T : \mathcal{V} \to \mathcal{W}$. Can we do better than just "set"? Yes!

Definition 9.4: Sums and Scalar Products of Linear Transformations

Given linear transformations $T_1, T_2 : \mathcal{V} \to \mathcal{W}$, for all $\vec{\boldsymbol{v}} \in \mathcal{V}$, $c \in F$, we have

- The sum $(T_1 + T_2)(\vec{\boldsymbol{v}}) = T_1(\vec{\boldsymbol{v}}) + T_2(\vec{\boldsymbol{v}})$
- The scalar product $(cT_1)(\vec{\boldsymbol{v}}) = cT_1(\vec{\boldsymbol{v}})$

As such, we have that $\mathcal{L}(\mathcal{V}, \mathcal{W}) \subset \mathcal{F}(\mathcal{V}, \mathcal{W})$. That is the set of all linear transformations is a subset of the set of all functions from $\mathcal{V}$ to $\mathcal{W}$.

Lemma 9.5

$\mathcal{L}(\mathcal{V}_F, \mathcal{W}_F)$ is a vector space over the field F.

Proof. For HW, we will show that fi T_1 and T_2 are linear, then $T_1 + T_2$ and cT_1 are also linear.

Further, we have the existence of a zero vector $\vec{\mathbf{0}}_{\mathcal{L}} = \prime : \mathcal{V} \to \mathcal{W}$, and the existence of an additive inverse $(-T)(\vec{\boldsymbol{v}}) = -(T(\vec{\boldsymbol{v}}))$.

Hence, all the definitions of a vector space hold. □

Even more, the matrices of T_1 and T_2 follow suit:

- $[T_1 + T_2]_\beta^\gamma = [T_1]_\beta^\gamma + [T_2]_\beta^\gamma$
- $[cT_1]_\beta^\gamma = c[T_1]_\beta^\gamma$

We can show the first point by looking at the definition of the matrix:

$$\begin{aligned}
[T_1 + T_2]_\beta^\gamma &= \begin{bmatrix} | & | & | \\ [(T_1 + T_2)(\vec{\boldsymbol{e}}_1)]_\gamma & \cdots & [(T_1 + T_2)(\vec{\boldsymbol{e}}_3)]_\gamma \\ | & | & | \end{bmatrix} \\
&= \begin{bmatrix} | & | & | \\ [T_1(\vec{\boldsymbol{e}}_1)]_\gamma & \cdots & [T_1(\vec{\boldsymbol{e}}_3)]_\gamma \\ | & | & | \end{bmatrix} + \begin{bmatrix} | & | & | \\ [T_2(\vec{\boldsymbol{e}}_1)]_\gamma & \cdots & [T_2(\vec{\boldsymbol{e}}_3)]_\gamma \\ | & | & | \end{bmatrix} \\
&= [T_1]_\beta^\gamma + [T_2]_\beta^\gamma
\end{aligned}$$

For homework, we will show the scalar multiplication case.

In summary, we have a one-to-one correspondence between the space of linear transformations and the space of $m \times n$ matrices:

$$\begin{array}{rcl} \mathcal{L}(\mathcal{V},\mathcal{W}) & \xleftrightarrow{\text{1-1 corr.}} & \mathcal{M}_{m\times n}(F) \\ T & \mapsto & \mathbf{M}_T = [T]_\beta^\gamma \\ T = L_{\mathbf{M}} & \leftarrow\!\!\shortmid & \mathbf{M} \\ [T(\vec{\boldsymbol{v}})]_\gamma = \mathbf{M}\cdot[\vec{\boldsymbol{v}}]_\beta & & \\ & & \\ T_1, T_2 & \mapsto & \mathbf{M}_1, \mathbf{M}_2 \\ T_1 + T_2 & \mapsto & \mathbf{M}_1 + \mathbf{M}_2 \\ cT_1 & \mapsto & c\mathbf{M}_1 \end{array}$$

9/17/2021

Lecture 10

Compositions of Transformations

10.1 Composition of Linear Transformations

Now, let us consider three linear transformations; $T_1 : \mathcal{V}_\beta^n \to \mathcal{W}_\gamma^m$ and $T_2 : \mathcal{W}_\gamma^m \to \mathcal{X}_\alpha^k$.

Can we find $T_3 = T_2 \circ T_1 : \mathcal{V} \to \mathcal{X}$? Is it linear? What is the matrix of $T_2 \circ T_1$?

Theorem 10.1: Compositions Preserve Linearity

$T_2 \circ T_1$ is linear.

Proof. We have that

$$\begin{aligned} (T_2 \circ T_1)(\vec{\boldsymbol{v}}_1 + \vec{\boldsymbol{v}}_2) &= T_2(T_1(\vec{\boldsymbol{v}}1 + \vec{\boldsymbol{v}}2)) \\ &= T_2(T_1(\vec{\boldsymbol{v}}_1) + T_1(\vec{\boldsymbol{v}}_2)) \\ &= T_2(T_1(\vec{\boldsymbol{v}}_1)) + T_2(T_1(\vec{\boldsymbol{v}})) \\ &= (T_2 \circ T_1)(\vec{\boldsymbol{v}}_1) + (T_2 \circ T_1)(\vec{\boldsymbol{v}}_2) \end{aligned}$$

For homework, you will show the preservation of scalar multiplication. □

Theorem 10.2: Matrix of Compositions

Given fixed bases β for $\mathcal{V}$, γ for $\mathcal{W}$, and α for $\mathcal{X}$, the matrix of composition $T_2 \circ T_1$ is the product of the matrices of T_2 and T_1:

$$[T_2 \circ T_1]_\beta^\alpha = [T_2]_\gamma^\alpha \cdot [T_1]_\beta^\gamma.$$

Corollary 10.3

$\mathcal{L}(\mathcal{V},\mathcal{W})$ and $\mathcal{M}_{m\times n}(F)$ correspond as vector spaces and have corresponding composition/multiplication between the various spaces. (This provides us with some extra structure on top of the vector spaces.)

If we have $\mathcal{L}(\mathcal{V},\mathcal{W})$ and $\mathcal{L}(\mathcal{W},\mathcal{X})$, and their corresponding matrices $\mathbf{M}_{m\times n}$ and $\mathbf{M}_{n\times k}$, we have

$$T_2 \circ T_1 \leftrightarrow \mathbf{M}_{T_2} \cdot \mathbf{M}_{T_1}.$$

In summary: the compositions of linear transformations correspond to multiplication of matrices. We have built extra structure on top of existing vector space structures.

We originally only had two operations under the vector space of matrices $\mathcal{M}(F)$ of various sizes; for vectors $\mathbf{A}$, $\mathbf{B}$, we had addition $\mathbf{A}+\mathbf{B}$ and $c\mathbf{A}$. Now, we've added a third operations: $\mathbf{A}_{m\times n}\cdot\mathbf{B}_{n\times k}=\mathbf{C}_{m\times k}$.

Recall that it was the composition of linear transformations $T_3 = T_2\circ T_1$ that inspired the definition of matrix multiplication $c_{ij}=\sum_{k=1}^n a_{ik}\cdot b_{kj}$.

10.2 Properties of Matrix Multiplication

Under the assumption that the matrices are of the right size,

- Distributivity: $\mathbf{A}(\mathbf{B}+\mathbf{C})=\mathbf{AB}+\mathbf{AC}$
- Scalar associativity: $c(\mathbf{AB})=(c\mathbf{A})\mathbf{B}=\mathbf{A}(c\mathbf{B})$

We could prove these on two levels; on homework, we will prove this on a level of matrices (that is, entry-wise):

1. Distributivity: $(\mathbf{A}(\mathbf{B}+\mathbf{C}))_{ij}=\sum_k a_{ik}(\mathbf{B}+\mathbf{C})_{kj}$
2. Scalar associativity: $(c(\mathbf{AB}))_{ij}=\sum_k ca_{ik}b_{kj}$

One thing that we know is (almost) never true is commutativity of matrices; $\mathbf{AB}\neq\mathbf{BA}$.

Another question we could ask is whether there exists a multiplicative identity:

$$\mathbf{I}_n=\begin{bmatrix}1 & \cdots & 0 & 0\\ \vdots & 1 & \ddots & 0\\ 0 & \ddots & 1 & \vdots\\ 0 & 0 & \cdots & 1\end{bmatrix}.$$

We'll use $\mathcal{I}$ to denote the corresponding identity transformation.

Definition 10.4: Identity Matrix

The identity matrix is defined as $(\mathbf{I}_n)_{ij}=\delta_{ij}=\begin{cases}1 & i=j\\ 0 & i\neq j\end{cases}$

We can also see that $\mathbf{A}_{m\times n}\cdot\mathbf{I}_n=\mathbf{A}_{m\times n}$ and $\mathbf{I}_{m\times m}\cdot\mathbf{A}_{m\times n}=\mathbf{A}_{m\times n}$.

Can we avoid brute force in proving the above? Yes; we can look at the linear transformations instead. We have $T:\mathcal{V}\to\mathcal{W}$ and $\mathcal{I}:\mathcal{W}\to\mathcal{X}$, and we want to consider $\mathcal{I}\circ T$.

We have $\mathcal{I}(\vec{\boldsymbol{w}})=\vec{\boldsymbol{w}}$, and we can show that $\mathcal{I}\circ T=T$, and similarly $T\circ\mathcal{I}=T$. Hence, matrices follow suit; $\mathbf{I}\cdot\mathbf{M}_T=\mathbf{M}_T\cdot\mathbf{I}=\mathbf{M}_T$.

Lastly, we can consider associativity:

$$\mathbf{A}_{m\times n}\cdot(\mathbf{B}_{n\times k})\cdot\mathbf{C}_{k\times l}=(\mathbf{A}_{m\times n}\cdot\mathbf{B}_{n\times k})\cdot\mathbf{C}_{k\times l}.$$

If we use linear transformations, we have $\mathcal{V}\xrightarrow{C}\mathcal{W}\xrightarrow{B}\mathcal{X}\xrightarrow{A}\mathcal{Y}$, We know that

$$A\circ(B\circ C)=(A\circ B)\circ C)=A\circ B\circ C.$$

This is because the composition of functions is associative (you can verify as well). As such, the matrices will follow suit.

9/20/2021

Lecture 11

Problem Solving with Compositions of Transformations, Introduction to Isomorphisms

Example 11.1

We can try seeing that linear transformations $T : \mathcal{V} \to \mathcal{V}$ satisfy $T^2 = T$.

We have $T^2 - TE = 0 = T^2 - ET$, or $T(T-E) = 0 = (T-E)T$.

If $\vec{\boldsymbol{v}} \in \operatorname{Ker} T$, then we'd have $T(\vec{\boldsymbol{v}}) = 0$ and we're done. If $\vec{\boldsymbol{v}} \in \operatorname{Im} T$, then let $\vec{\boldsymbol{w}} = T(\vec{\boldsymbol{v}})$ for some $\vec{\boldsymbol{w}} \in \mathcal{V}$. We have $(T-E)\vec{\boldsymbol{w}} = \vec{\boldsymbol{0}}$ and $T(\vec{\boldsymbol{w}}) = \vec{\boldsymbol{w}}$.

11.1 Isomorphisms

Definition 11.2: Isomorphism

$T : \mathcal{V} \to \mathcal{W}$ is an isomorphism if and only if T is one-to-one and onto. We denote this as $\mathcal{V} \cong \mathcal{W}$.

Note: we can label a linear transformation an *isomorphism*, and two vector spaces to be *isomorphic*.

Recall that T is one-to-one iff $\operatorname{Ker} T = \{\vec{\boldsymbol{0}}\}$ and T is onto iff $\operatorname{Im} T = \mathcal{W}$.

To be an isomorphism a linear transformation needs to be (one-to-one) injective and (onto) surjective, meaning it is bijective (a one-to-one correspondence).

What does it mean to be bijective? T would need to be invertible. That is, if we have a linear transformation T, there exists a T^{-1} that gives us the pre-images.

More formally, we have $T \circ T^{-1} = \mathcal{I}_{\mathcal{V}}$ and $T^{-1} \circ T = \mathcal{I}_{\mathcal{W}}$. Here, we denote T^{-1} to be the inverse of a bijective T.

One question you could ask is whether T^{-1} is linear;

Proof. Suppose we have $\vec{\boldsymbol{w}}_1, \vec{\boldsymbol{w}}_1 \in \mathcal{W}$, where $\vec{\boldsymbol{v}}_1 = T^{-1}(\vec{\boldsymbol{w}}_1)$ and $\vec{\boldsymbol{v}}_2 = T^{-1}(\vec{\boldsymbol{w}}_2)$.

We want to show whether $T^{-1}(\vec{\boldsymbol{w}}_1 + \vec{\boldsymbol{w}}_2) = T^{-1}(\vec{\boldsymbol{w}}_1) + T^{-1}(\vec{\boldsymbol{w}}_2)$

Since T and T^{-1} are inverses, we have $T(\vec{\boldsymbol{v}}_1) = \vec{\boldsymbol{w}}_1$ and $T(\vec{\boldsymbol{v}}_2) = \vec{\boldsymbol{w}}_2$. This means that we can apply linearity of T to get that $\vec{\boldsymbol{w}}_1 + \vec{\boldsymbol{w}}_2 = T(\vec{\boldsymbol{v}}_1) + T(\vec{\boldsymbol{v}}_2) = T(\vec{\boldsymbol{v}}_1 + \vec{\boldsymbol{v}}_2)$.

We can then invert—we have $T^{-1}(\vec{\boldsymbol{w}}_1 + \vec{\boldsymbol{w}}_2) = \vec{\boldsymbol{v}}_1 + \vec{\boldsymbol{v}}_2 = T^{-1}(\vec{\boldsymbol{w}}_1) + T^{-1}(\vec{\boldsymbol{w}}_2)$

Hence, T^{-1} preserves sums. For HW, we will show that $T^{-1}(c\vec{\boldsymbol{w}}_1) = cT^{-1}(\vec{\boldsymbol{w}}_1), \forall c \in F$. □

A next question could be: what happens to bases under isomorphisms?

Theorem 11.3: Isomorphisms Mapping Bases

If the linear transformation $T : \mathcal{V} \to \mathcal{W}$ is one-to-one and onto, then any basis of $\mathcal{V}$ maps to a basis of $\mathcal{W}$ by T.

Proof. If T is one-to-one (injective), then linearly independent sets get mapped to linearly independent sets by T.

If T is onto (surjective), then spanning sets get mapped to spanning sets by T.

Therefore, we can put all this together and show that any basis gets mapped to another basis by T. □

Theorem 11.4: Onto Mapping Spanning Sets

A spanning set gets mapped to a spanning set by T if T is onto.

Proof. Let $\operatorname{span}\{\vec{v}_1, \dots, \vec{v}_n\} = \mathcal{V}$, and let $\vec{w}_i = T(\vec{v}_i)$ for $i = 1, 2, \dots, n$.

To show that $\vec{w}_i$ span $\mathcal{W}$, let $\vec{w} \in \mathcal{W}$. Since T is onto, we must have $\vec{w} = T(\vec{v})$ for some $\vec{v} \in \mathcal{V}$, and $\vec{v}$ must be able to be written as a linear combination of the basis vectors. That is, $\vec{v} = \sum_i c_i \vec{v}_i$ for $c_i \in F$.

Applying both sides with T, we have $\vec{w} = T(\vec{v}) = \sum_{i=1}^n c_i T(\vec{v}_1) = \sum_{i=1}^n c_i \vec{w}_i$

This means that any vector $\vec{w} \in \mathcal{W}$ can be written as a linear combination of $\vec{w}_i$, and as such

$$\operatorname{span}\{\vec{w}_1, \dots, \vec{w}_n\} = \mathcal{W}.$$

□

Theorem 11.5: Ismorphism Dimensions

If $T : \mathcal{V} \to \mathcal{W}$ is an isomorphism, then $\dim \mathcal{V} = \dim \mathcal{W}$.

Proof. In brief, we've just shown that bases of $\mathcal{V}$ will be mapped to bases of $\mathcal{W}$, and thus the two bases must be of the same dimension. □

9/22/2021

Lecture 12

Properties of Isomorphisms, Invertibility

Continuing on from last time, we can look at the converse.

Theorem 12.1: Isomorphism Dimensions (Converse)

If $\dim \mathcal{V} = \dim \mathcal{W} (< \infty)$, then there exists some linear transformation $T : \mathcal{V} \to \mathcal{W}$ that is an isomorphism.

Proof. Suppose we have a (fixed) basis $\beta = \{\vec{v}_1, \dots, \vec{v}_n\}$ for $\mathcal{V}$ and a (fixed) basis $\gamma = \{\vec{w}_1, \dots, \vec{w}_n\}$ for $\mathcal{W}$.

Suppose we create a linear function $T : \mathcal{V} \to \mathcal{W}$ such that $T(\vec{v}_i) = \vec{w}_i$ for all i.

Since we know what T does to the bases of $\mathcal{V}$, this uniquely defines the linear transformation, and we can extend this to any vector $\vec{v}$; that is,

$$T(\vec{v}) = T(c_1 \vec{v}_1 + c_2 \vec{v}_2 + \cdots + c_n \vec{v}_n) = c_1 \vec{w}_1 + \cdots + c_n \vec{w}_n.$$

Now, we need to prove that T is an isomorphism (that is, we also need to show that T is bijective; we leave the proof that T is linear on HW).

What is $[T]_\beta^\gamma$? It's the identity $\mathbf{I}_n$ (because we've defined this transformation to map the basis β to the basis γ).

From here, we can directly construct the inverse $T^{-1} : \mathcal{W} \to \mathcal{V}$, in a very similar way. That is, we let $U(\vec{w}_i) = \vec{v}_i$, for all i, extending this linearly to all $\vec{w} \in \mathcal{W}$.

We can check that T and U are indeed inverses of each other; we can see that $T \circ U$ fixes the basis, and is linear (because the composition of linear transformations is also linear). This means that $T \circ U$ fixes every single vector in $\mathcal{W}$. Similar reasoning can be applied to $U \circ T$, and as such $U = T^{-1}$.

Therefore, T is an isomorphism. $\square$

Definition 12.2: Invertible Linear Transformations

Let $T : \mathcal{V} \to \mathcal{W}$ be a linear transformation. T is invertible if there exists an inverse $T^{-1} : \mathcal{W} \to \mathcal{V}$ such that $T \circ T^{-1} = \mathcal{I}_{\mathcal{W}}$, and $T^{-1} \circ T = \mathcal{I}_{\mathcal{V}}$.

Further, the linear transformation T is invertible if and only if T is one-to-one and onto (T is bijective).

Theorem 12.3: Isomorphic Vector Spaces

Two vector spaces are isomorphic (that is, $\mathcal{V} \cong \mathcal{W}$) if and only if $\dim \mathcal{V} = \dim \mathcal{W}$ (for $\mathcal{V}$ and $\mathcal{W}$ of finite dimension).

12.1 Equivalence Relations

An equivalence relation on vector spaces over a field F has three properties:

- (Reflexivity) $\mathcal{V} \cong \mathcal{V}$ (ex. one such isomorphism is $\mathcal{I}_{\mathcal{V}}$).
- (Symmetric) $\mathcal{V} \cong \mathcal{W}$ with the isomorphism T implies that $\mathcal{W} \cong \mathcal{V}$ with the isomorphism T^{-1}.
- (Transitivity) $\mathcal{V} \cong \mathcal{W}$ by T_1 and $\mathcal{W} \cong \mathcal{U}$ by T_2 implies that $\mathcal{V} \cong \mathcal{U}$ by $T_2 \circ T_1$.

As a consequence, all vector spaces over the field F fall into disjoint "bags" (classes) of isomorphic spaces. By what criteria have we grouped these spaces? By dimension.

This is because of the theorem we just showed previously for finite dimension spaces (and it turns out this also happens for spaces of infinite dimension).

Specifically each class of vector spaces is defined to be the set of all vector spaces with dimension n. The idea here is that we can choose one representative from each bag; whatever we say about this representative vector space will always apply to other vectors in this (infinite) class.

12.2 Isomorphisms and Matrices

Let $\mathcal{V}^n_\beta \cong \mathcal{W}^n_\gamma$ through an isomorphism T. Suppose we have $[T]^\gamma_\beta = \mathbf{A}_{n\times n}$ and $[T^{-1}]^\beta_\gamma = \mathbf{B}_{n\times n}$.

We know that $T^{-1} \circ T = \mathcal{I}_{\mathcal{V}} : \mathcal{V} \to \mathcal{V}$, so this means that $[T^{-1} \circ T]^\beta_\beta = [\mathcal{I}_{\mathcal{V}}]^\beta_\beta = \mathbf{I}_n = \mathbf{BA}$.

We further have that $T \circ T^{-1} = \mathcal{I}_{\mathcal{W}} : \mathcal{W} \to \mathcal{W}$, so thi smeans that $[T \circ T^{-1}]^\gamma_\gamma = [\mathcal{I}_{\mathcal{W}}]^\gamma_\gamma = \mathbf{I}_n = \mathbf{AB}$.

Definition 12.4: Invertible Matrices

Let $\mathbf{A} \in \mathcal{M}_{n\times n}(F)$. Then $\mathbf{A}$ is invertible if $m = n$ and there exists a $\mathbf{B}_{n\times n}$ such that $\mathbf{AB} = \mathbf{I}_n$ and $\mathbf{BA} = \mathbf{I}_n$.

Is $\mathbf{B}$ unique? Suppose we have two inverses $\mathbf{B}$ and $\mathbf{B}_1$. We know that $\mathbf{AB} = \mathbf{I}_n = \mathbf{AB}_1$ and $\mathbf{BA} = \mathbf{I}_n = \mathbf{B}_1\mathbf{A}$. We want to show that $\mathbf{B} = \mathbf{B}_1$:

$$\mathbf{B} = \mathbf{I}_n\mathbf{B} = (\mathbf{B}_1\mathbf{A})\mathbf{B} = \mathbf{B}_1(\mathbf{AB}) = \mathbf{B}_1\mathbf{I}_n = \mathbf{B}_1.$$

Moreover, if a matrix is a one-sided inverse of $\mathbf{A}_{n\times n}$, the matrix is automatically a double-sided inverse of $\mathbf{A}$.

Theorem 12.5: Commutativity of Inverse Matrix

Suppose $\mathbf{A}, \mathbf{B} \in \mathcal{W}_{n\times n}(F)$. If $\mathbf{AB} = \mathbf{I}_n$, then $\mathbf{BA} = \mathbf{I}_n$.

Proof. Let $T: \mathcal{F}^n \to \mathcal{F}^n$ be linear, where $T(\vec{\boldsymbol{v}}) = \mathbf{A}\vec{\boldsymbol{v}}$, and $U: \mathcal{F}^n \to \mathcal{F}^n$ be linear, where $U(\vec{\boldsymbol{v}}) = \mathbf{B}\vec{\boldsymbol{v}}$.

If we look at the composition $T \circ U$, we have

$$(T \circ U)(\vec{\boldsymbol{v}}) = \mathbf{AB}\vec{\boldsymbol{v}} = \mathbf{I}_n\vec{\boldsymbol{v}} = \vec{\boldsymbol{v}}.$$

This means that $T \circ U = \mathcal{I}_{\mathcal{F}^n}$. However, from here, we can conclude that U is one-to-one: by contradiction, suppose U is not one-to-one. This means that $\vec{\boldsymbol{x}}_1$ and $\vec{\boldsymbol{x}}_2$ both map to $\vec{\boldsymbol{y}}$. However, this means that T is undefined—it cannot send $\vec{\boldsymbol{y}}$ to both $\vec{\boldsymbol{x}}_1$ and $\vec{\boldsymbol{x}}2$.

As such, we can conclude that $\operatorname{Ker} U = \{\vec{\mathbf{0}}\}$, and U is one-to-one. Further, by the dimension theorem, we have $n = \operatorname{null}(U) + \operatorname{rank}(U)$, meaning $\operatorname{rank}(U) = n$, leading to the fact that U is onto.

This means that U is an isomorphism, meaning T is also an isomorphism. □

12.3 Criteria for Isomorphisms

Theorem 12.6: Criteria for Isomorphisms

Let $\dim V = \dim W = n$ and $T: \mathcal{V} \to \mathcal{W}$ be linear. Then, the following are equivalent:

$$T \text{ is an isomorphism} \iff T \text{ is one-to-one} \iff T \text{ is onto}$$

Proof. We already have that if T is an isomorphism then T is one-to-one and onto. We further have that by the dimension theorem if T is one-to-one, then T is onto as well. Together, this implies that T is an isomorphism. □

More generally,

Corollary 12.7

If $T: \mathcal{V} \to \mathcal{W}$ is linear, then the following are equivalent:

- T is one-to-one and onto
- T is an isomorphism
- T has an inverse $T^{-1}: \mathcal{W} \to \mathcal{V}$
- For any bases β of $\mathcal{V}$ and γ of $\mathcal{W}$, $[T]_\beta^\gamma$ is invertible

12.4 Classification

Firstly, we can see that $\mathcal{V}_F \cong \mathcal{F}^n$ if and only if $\dim \mathcal{V} = n$.

Definition 12.8: Standard Isomorphism

The *standard isomorphism* for $\mathcal{V}$ with respect to basis $\beta = \{\vec{\boldsymbol{v}}_1, \ldots, \vec{\boldsymbol{v}}_n\}$ is the coordinate isomorphism. That is,

$$\phi_\beta : \mathcal{V} \to \mathcal{F}^n := \vec{\boldsymbol{v}} \mapsto [\vec{\boldsymbol{v}}]_\beta = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}.$$

A corollary of this is that all properties of $\mathcal{F}^n$ can be ascribed to all vector spaces $\mathcal{V} \cong \mathcal{F}^n$ through this isomorphism.

9/24/2021

Lecture 13

Change of Bases

How do we write a vector from standard basis to a different basis? Suppose we take $\mathbb{R}^2$ and the basis $\beta = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right\}$.

How do we write $\begin{bmatrix} a \\ b \end{bmatrix}_e$ as $\begin{bmatrix} x \\ y \end{bmatrix}_\beta$? If we write out the system, we have

$$a \begin{bmatrix} 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \end{bmatrix} = x \begin{bmatrix} 1 \\ 1 \end{bmatrix} + y \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

This implies that we have

$$\begin{aligned} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} &= \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ \begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}^{-1} \begin{bmatrix} a \\ b \end{bmatrix} \end{aligned}$$

If we call this RHS matrix (on the first line) $\mathbf{R}$, we can see that $\mathbf{R}$ expresses the old basis β in terms of the new basis e; the inverse $\mathbf{R}^{-1}$ expresses the old basis e in terms of the new basis β.

That is, the columns of $\mathbf{R}$ are $[\vec{\boldsymbol{v}}_1]_e$ and $[\vec{\boldsymbol{v}}_2]_e$, whereas the columns of $\mathbf{R}^{-1}$ are $[\vec{\boldsymbol{e}}_1]_\beta$ and $[\vec{\boldsymbol{e}}_2]_\beta$.

In summary, suppose we have two bases $\gamma_1 = \{\vec{\boldsymbol{v}}_1, \dots, \vec{\boldsymbol{v}}_n\}$ and $\gamma_2 = \{\vec{\boldsymbol{u}}_1, \dots, \vec{\boldsymbol{u}}_n\}$ for $\mathcal{V}$, where $\mathcal{V}_F$ is of dimension n.

Let $\mathbf{Q}$ be the matrix expressing the old basis γ_1 in terms of the new basis γ_2. This matrix is

$$\mathbf{Q} = \begin{bmatrix} | & | & & | \\ [\vec{\boldsymbol{v}}_1]_{\gamma_2} & [\vec{\boldsymbol{v}}_2]_{\gamma_2} & \cdots & [\vec{\boldsymbol{v}}_n]_{\gamma_2} \\ | & | & & | \end{bmatrix}.$$

Theorem 13.1: Change of Coordinates from γ_1 to γ_2

The coordinates in the new basis γ_2 are obtained from the old basis γ_1 by:

$$[\vec{\boldsymbol{v}}]_{\gamma_2} = \mathbf{Q}[\vec{\boldsymbol{v}}]_{\gamma_1} = [\mathcal{L}_{\mathcal{V}}]_{\gamma_1}^{\gamma_2} \cdot [\vec{\boldsymbol{v}}]_{\gamma_1}.$$

The key idea here is that $\mathbf{Q}$ is the matrix of the linear transformation $T : \mathcal{V} \to \mathcal{V}$ sending $\vec{\boldsymbol{v}}_1 \to \vec{\boldsymbol{v}}_1, \dots, \vec{\boldsymbol{v}}_n \to \vec{\boldsymbol{v}}_n$, but written in the bases $\gamma_1 \to \gamma_2$.

Definition 13.2: Change of Coordinate Matrix

The matrix $\mathbf{Q} = [\mathcal{I}_{\mathcal{V}}]_{\gamma_1}^{\gamma_2}$ is called the change of coordinate matrix, where $\mathbf{Q}^{-1} = [\mathcal{I}_{\mathcal{V}}]_{\gamma_2}^{\gamma_1}$.

Note that both $\mathbf{Q}$ and $\mathbf{Q}^{-1}$ are invertible.

Similarly, we have $[\vec{\boldsymbol{v}}]_{\gamma_1} = \mathbf{Q}^{-1}[\vec{\boldsymbol{v}}]_{\gamma_2} = [\mathcal{I}_{\mathcal{V}}]_{\gamma_2}^{\gamma_1} \cdot [\vec{\boldsymbol{v}}]_{\gamma_2}$

Example 13.3

Suppose we have the basis

$$\beta = \left\{ \begin{bmatrix} -4 \\ 3 \end{bmatrix}, \begin{bmatrix} 8 \\ -9 \end{bmatrix} \right\}$$
$$\beta' = \left\{ \begin{bmatrix} 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 4 \\ 0 \end{bmatrix} \right\}$$

Here, we let $\beta = \{\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2\}$ and $\beta' = \{\vec{\boldsymbol{u}}_1, \vec{\boldsymbol{u}}_2\}$.

If $\vec{\boldsymbol{v}} = \begin{bmatrix} 1 \\ 6 \end{bmatrix}_{\beta'}$, what is $[\vec{\boldsymbol{v}}]_\beta$?

We can see that

$$\mathbf{Q} = [\mathcal{I}_\mathcal{V}]_{\beta'}^{\beta} = \begin{bmatrix} | & | \\ [\vec{\boldsymbol{u}}_1]_\beta & [\vec{\boldsymbol{u}}_2]_\beta \\ | & | \end{bmatrix}.$$

This isn't a very nice calculation, so we instead can find

$$\mathbf{Q}^{-1} = [\mathcal{I}_\mathcal{V}]_{\beta}^{\beta'} = \begin{bmatrix} | & | \\ [\vec{\boldsymbol{v}}_1]_{\beta'} & [\vec{\boldsymbol{v}}_2]_{\beta'} \\ | & | \end{bmatrix} = \begin{bmatrix} 1 & -3 \\ -1 & 2 \end{bmatrix}.$$

We can also see that $\mathbf{Q} = -\begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix}$.

Hence, we have

$$[\vec{\boldsymbol{v}}]_\beta = [\mathcal{I}_\mathcal{V}]_{\beta'}^{\beta} [\vec{\boldsymbol{v}}]_{\beta'} = \begin{bmatrix} -2 & -3 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 6 \end{bmatrix} = \begin{bmatrix} -20 \\ -7 \end{bmatrix}_\beta.$$

We could further check that this calculation was right by converting back to β'.

Definition 13.4: Linear Operator

If we have a linear transformation $T : \mathcal{V} \to \mathcal{V}$, we call T a *linear operator* on $\mathcal{V}$.

Suppose we have a linear operator $T : \mathcal{V} \to \mathcal{V}$, and we have $\gamma_1 = \{\vec{\boldsymbol{v}}_1, \dots, \vec{\boldsymbol{v}}_n\}$ and $\gamma_2 = \{\vec{\boldsymbol{u}}_1, \dots, \vec{\boldsymbol{u}}_n\}$ as two bases of $\mathcal{V}$. We know $[T]_{\gamma_1}^{\gamma_1} = [T]_{\gamma_1}$, and we want to find $[T]_{\gamma_2}^{\gamma_2} = [T]_{\gamma_2}$.

Theorem 13.5: Change of Bases

$$[T]_{\gamma_2}^{\gamma_2} = [\mathcal{I}_\mathcal{V}]_{\gamma_1}^{\gamma_2} \cdot [T]_{\gamma_1}^{\gamma_1} \cdot [\mathcal{I}_\mathcal{V}]_{\gamma_2}^{\gamma_1}.$$

Proof. We can view this as a triple composition of linear transformations. We have $[\mathcal{I}_\mathcal{V}]_{\gamma_1}^{\gamma_2} \cdot [T]_{\gamma_1}^{\gamma_1}$ as the composition $\mathcal{I}_\mathcal{V} \circ T$, from $\mathcal{V}_{\gamma_1} \xrightarrow{T} \mathcal{V}_{\gamma_1} \xrightarrow{\mathcal{I}_\mathcal{V}} \mathcal{V}_{\gamma_2}$. This is just $[T]_{\gamma_1}^{\gamma_2}$.

We can follow the same idea, simplifying $[T]_{\gamma_1}^{\gamma_2} \cdot [\mathcal{I}_\mathcal{V}]_{\gamma_2}^{\gamma_1} = [T \circ \mathcal{I}_\mathcal{V}]_{\gamma_2}^{\gamma_2} = [T]_{\gamma_2}^{\gamma_2}$. □

Example 13.6

Suppose we have a linear operator $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $[T]_e = \begin{bmatrix} -1 & 2 \\ -3 & 4 \end{bmatrix}$.

We have $\beta = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}_e, \begin{bmatrix} 2 \\ 3 \end{bmatrix}_e \right\}$. What is $[T]_\beta$?

We already have the new basis expressed as the standard basis (this is just $[T]_e$, as the columns are $[\vec{\boldsymbol{v}}_i]_e$).

This means that $[T]_e = [\mathcal{I}_\mathcal{V}]^e_\beta$. The inverse is then $\begin{bmatrix} 3 & -2 \\ -2 & 1 \end{bmatrix} = [\mathcal{I}_\mathcal{V}]^\beta_e$.

Remember that we're given $[T]^e_e$, and want to determine $[T]^\beta_\beta$. This means that we have

$$[T]_\beta = [\mathcal{I}_\mathcal{V}]^\beta_e [T]_e [\mathcal{I}_\mathcal{V}]^e_\beta = \begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}.$$

Example 13.7: Reflection Across a Line

Find the matrix of reflection in $\mathbb{R}^2$ across line ℓ (described by $y = \frac{4}{3}x$) in a suitable basis.

If we choose the basis where ℓ is the x-axis, along with the second basis vector perpendicular to ℓ, we have the bases $\beta = \left\{ \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \begin{bmatrix} -4 \\ 3 \end{bmatrix} \right\}$.

We then have

$$[T]_\beta = \begin{bmatrix} | & | \\ [\vec{\boldsymbol{v}}_1]_\beta & [-\vec{\boldsymbol{v}}_2]_\beta \\ | & | \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

We want the matrix $[T]_e$, so we have

$$[T]_e = [\mathcal{I}_{\mathbb{R}^2}]^e_\beta [T]_\beta [\mathcal{I}_{\mathbb{R}^2}]^\beta_e = \begin{bmatrix} 3 & -4 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \frac{1}{25} \begin{bmatrix} 3 & 4 \\ -4 & 3 \end{bmatrix} = \frac{1}{25} \begin{bmatrix} -7 & 24 \\ 24 & 7 \end{bmatrix}.$$

In conclusion, the standard unit basis is not always the most convenient one!

9/27/2021

Lecture 14

Similar Matrices, Dual Bases

Definition 14.1: Similar Matrices

If $\mathbf{A}, \mathbf{B} \in \mathcal{M}_{n\times n}(F)$ and there exists an invertible matrix $\mathbf{Q} \in \mathcal{M}_{n\times n}(F)$ such that $\mathbf{A} = \mathbf{Q}^{-1}\mathbf{B}\mathbf{Q}$ then $\mathbf{A}$ is similar to $\mathbf{B}$; we denote this as $\mathbf{A} \sim \mathbf{B}$.

Theorem 14.2: Matrices of Linear Transformations are Similar

Suppose we have a linear $T\colon \mathcal{V} \to \mathcal{V}$. Then all matrices of T are similar. Conversely, if $\mathbf{A} \sim \mathbf{B}$, then $\mathbf{A}$ and $\mathbf{B}$ are matrices of the same linear transformation $T\colon \mathcal{V} \to \mathcal{V}$.

Proof. ($\implies$) We have $\mathbf{A} = \mathbf{Q}^{-1}\mathbf{B}\mathbf{Q}$ where $\mathbf{Q} = [\mathcal{I}_{\mathcal{V}}]_\beta^\gamma$.

($\impliedby$) If $\mathbf{A} = \mathbf{Q}^{-1}\mathbf{B}\mathbf{Q}$, then let $T : \mathcal{F}^n \to \mathcal{F}^n$ such that $T(\vec{\boldsymbol{v}}) = \mathbf{B}\vec{\boldsymbol{v}}_e$. That is, $[T]_e = \mathbf{B}$.

Suppose we have

$$\mathbf{Q} = \begin{bmatrix} | & & | \\ [\vec{\boldsymbol{v}}_1]_e & \cdots & [\vec{\boldsymbol{v}}_n]_e \\ | & & | \end{bmatrix}.$$

We know the columns are linearly independent because it is invertible, so as such $\beta = \{\vec{\boldsymbol{v}}_1, \cdots, \vec{\boldsymbol{v}}_n\}$ be a basis for $\mathcal{V}$ (as $\mathcal{V}$ has dimension n).

Hence, this means that $\mathbf{Q} = [\mathcal{I}_{\mathcal{V}}]_\beta^e$ and $\mathbf{Q}^{-1} = [\mathcal{I}_{\mathcal{V}}]_e^\beta$.

We then have $\mathbf{A} = \mathbf{Q}^{-1}\mathbf{B}\mathbf{Q} = [\mathcal{I}_{\mathcal{V}}]_e^\beta [T]_e^e [\mathcal{I}_{\mathcal{V}}]_\beta^e = [T]_\beta^\beta$. ☐

14.1 Properties of Similar Matrices

Recall that isomorphisms are an equivalence relation on vector spaces. Is there such a relation on square $n \times n$ matrices? Yes—similarity.

For homework, you will need to show that $\mathbf{A}\,\mathbf{A}$, if $\mathbf{A} \sim \mathbf{B}$ (through $\mathbf{Q}$) then $\mathbf{B} \sim \mathbf{A}$ (through $\mathbf{Q}^{-1}$), and finally if $\mathbf{A} \sim \mathbf{B}$ (through $\mathbf{Q}_1$) and $\mathbf{B} \sim \mathbf{C}$ (through $\mathbf{Q}_2$) then $\mathbf{A} \sim \mathbf{C}$ (through $\mathbf{Q}_2\mathbf{Q}_1$).

Lemma 14.3

If $\mathbf{A} \sim \mathbf{B}$ then $\operatorname{tr}\mathbf{A} = \operatorname{tr}\mathbf{B}$.

Proof. Recall that $\operatorname{tr}(\mathbf{XY}) = \operatorname{tr}(\mathbf{YX})$ for all square $n \times n$ matrices.

Further, we ahve

$$\mathbf{A} = \mathbf{Q}^{-1}\mathbf{B}\mathbf{Q} \implies \operatorname{tr}\mathbf{A} = \operatorname{tr}\Big(\overbrace{\mathbf{Q}^{-1}}^{\mathbf{X}}\overbrace{\mathbf{B}\mathbf{Q}}^{\mathbf{Y}}\Big) = \operatorname{tr}\big(\mathbf{B}\big(\mathbf{Q}\mathbf{Q}^{-1}\big)\big) = \operatorname{tr}\mathbf{B}.$$

☐

14.2 Dual Spaces

(We'll only be talking about the bare-bones of dual spaces.)

Suppose we start with a vector space $\mathcal{V}$ over F. If we take $\mathcal{L}(\mathcal{V}, F)$, the set of all linear transformations from $\mathcal{V} \to F$, these are the set of *linear functionals* on $\mathcal{V}$.

Example 14.4

The dual of $\mathbb{R}^2$ is $\mathcal{L}(\mathbb{R}^2, \mathbb{R})$.

For example

$$\begin{bmatrix} a \\ b \end{bmatrix}_e \mapsto a + b.$$

This is a functional on $\mathbb{R}^2$.

Example 14.5

The dual of $\mathcal{V} = \mathcal{M}_{n\times n}(\mathbb{C}) \to \mathbb{C}$.

FOr example,

$$f(\mathbf{A}) = \operatorname{tr}\mathbf{A}.$$

This is a functional on $\mathcal{M}_{n\times n}$.

Suppose we think about a functional $f : \mathcal{V} \to F$. Our inputs are vectors, and our outputs are scalars; we can think of f as a (linear) multivariable function.

Definition 14.6: Dual Space

Suppose we have the vector space $\mathcal{V}_F$. Then, $\mathcal{L}(\mathcal{V}, F) = \mathcal{V}^*$ is the dual space of $\mathcal{V}$.

This is also a vector space but consists of functionals on $\mathcal{V}$.

If $\dim \mathcal{V} = n$, then

$$\dim \mathcal{V}^* = \dim \mathcal{L}(\mathcal{V}, F) = \dim \mathcal{V} \cdot \dim F = n \cdot 1 = n.$$

In conclusion, we must have that $\mathcal{V} \cong \mathcal{V}^*$. (The reason why $\dim \mathcal{L}(\mathcal{V}, F) = \dim \mathcal{V} \cdot \dim F$ is because we have an isomorphism between $\mathcal{M}_{m\times n}$, constructed by the bases of $\mathcal{V}$ and F).

14.2.1 Bases for Dual Spaces

Let us define $f_i : \mathcal{V} \to F$, and write

$$[\vec{\boldsymbol{v}}]_\beta = \begin{bmatrix} a_1 \\ \cdots \\ a_i \\ \cdots \\ a_n \end{bmatrix} \mapsto a_i.$$

That is, this function is the ith coordinate. We claim that this is another basis.

Let $\beta = \{\vec{\boldsymbol{v}}_1, \ldots, \vec{\boldsymbol{v}}_n\}$ be a basis for $\mathcal{V}$. Then, we have $f_i(\vec{\boldsymbol{v}}_j) = \delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$.

Lemma 14.7

If $\beta = \{\vec{\boldsymbol{v}}_1, \ldots, \vec{\boldsymbol{v}}_n\}$ is a basis for $\mathcal{V}$, then $\gamma = \{f_1, \ldots, f_n\}$ is a basis for $\mathcal{V}^*$ called the dual basis to β.

Moreover, any $f \in \mathcal{V}^*$ is given as the unique linear combination

$$f = f(\vec{\boldsymbol{v}}_1) f_1 + f(\vec{\boldsymbol{v}}_2) f_2 + \cdots + f(\vec{\boldsymbol{v}}_n) f_n.$$

Proof. It suffices to show that the latter is true, because it means that f_i's span $\mathcal{V}^*$, and since $\mathcal{V}^*$ are of dimension n, they must be bases for $\mathcal{V}^*$.

All we have to check is that the two sides ($\in \mathcal{L}(\mathcal{V}, F)$) agree on a basis of $\mathcal{V}$; say, on β:

$$f(\vec{\boldsymbol{v}}_1) = f(\vec{\boldsymbol{v}}_1)\overset{f_1(\vec{\boldsymbol{v}}_2)}{1} + \cancel{f(\vec{\boldsymbol{v}}_2) f_2(\vec{\boldsymbol{v}}_1)} + \cdots + \cancel{f(\vec{\boldsymbol{v}}_n) f_n(\vec{\boldsymbol{v}}_1)}.$$

As such, $\gamma = \{f_1, \ldots, f_n\}$ is in fact a basis for the dual $\mathcal{V}^*$. ☐

Example 14.8

Let $\mathcal{V} = \mathbb{R}^2$, where $\mathcal{V}^* = \mathcal{L}(\mathbb{R}^2, \mathbb{R})$, and

$$\beta = \{\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\2\end{bmatrix}\} \qquad \text{basis for } \mathcal{V}$$
$$\beta^* = \{f_1, f_2\} \qquad \text{dual basis of } \beta$$

Suppose we start with a vector $\begin{bmatrix}a\\b\end{bmatrix}_\beta$. We have ... **finish**

If we have $\mathcal{V}^n_{beat} \to \mathcal{W}^m_\gamma$ through T, we can take a functional $F : \mathcal{W} \to F$. This means that $f \circ T = g \ni \mathcal{L}(\mathcal{V}, F)$. We define this $T^* = f \circ T$.

On the level of matrices, suppose we have $[T]^\gamma_\beta = \mathbf{A} \in \mathcal{M}_{m\times n}(F)$. We want a linear transformation corresponding to $\mathbf{A}$, such that $T^* : \mathcal{W}^* \to \mathcal{V}^*$. It turns out that $[T^*]^{\beta^*}_{\gamma^*} = \mathbf{A}^T \in \mathcal{M}_{n\times m}$.

Theorem 14.9: Dual Space Isomorphisms

If $\dim \mathcal{V} = n$, then $\mathcal{V}$, $\mathcal{V}^*$, $(\mathcal{V}^*)^*$ are all of dimension n, and hence all isomorphic!

Further there does not exist any natural isomorphism between $\mathcal{V}$ and $\mathcal{V}^*$ (because it always depends on the choice of bases).

However, there is a natural (canonical isomorphism) $\mu : \mathcal{V} \to \mathcal{V}^{**}$.

10/1/2021

Lecture 15

Systems of Linear Equations—Theoretical Aspects

Definition 15.1: System of Linear Equations

A system of m linear equations with n unknowns $x_1, \ldots, x_n$ over the field F is

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

We can rewrite this as

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m_1} & a_{m_2} & \cdots & a_{mn} \end{bmatrix}.$$

This is the coefficient matrix of the system of equations; we have

$$\mathbf{A}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

That is, $\mathbf{A}\vec{x} = \vec{b}$, where $\vec{x} \in \mathcal{F}^n$ and $\vec{b} \in \mathcal{F}^m$.

A solution to a system of equations is any vector $\vec{x} \in \mathcal{F}^n$ such that $\mathbf{A}\vec{x} = \vec{b}$.

Definition 15.2: Homogeneous and Non-Homogeneous Systems

Suppose we have a system of linear equations $\mathbf{A}\vec{x} = \vec{b}$. If $\vec{b} = \vec{0}$, then the system is homogeneous. If $\vec{b} \neq \vec{0}$, then the system is non-homogeneous.

Example 15.3

Suppose we have the non-homogeneous system

$$\begin{cases} 2x_1 + 3x_2 = 5 \\ -x_1 - 3x_2 = 0 \end{cases}.$$

This corresponds to the matrix equation $\begin{bmatrix} 2 & 3 \\ -1 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$.

How do we solve such systems of linear equations?

We're usually given a non-homogeneous system $\mathbf{A}\vec{x} = \vec{b}$, but we'll first solve the homogeneous system $\mathbf{A}\vec{x} = \vec{0}$, then return to $\mathbf{A}\vec{x} = \vec{b}$ and produce one non-homogeneous solution $\vec{x}_0$.

Using the last two steps, we can conclude that the solution set to $\mathbf{A}\vec{x} = \vec{b}$ is $\{\vec{x}_0 + \vec{x} : \mathbf{A}\vec{x} = \vec{0},\ \mathbf{A}\vec{x}_0 = \vec{b}\}$.

15.1 Solving homogeneous systems

Lemma 15.4

Let $\mathbf{A}\vec{x} = \vec{0}$ be a homogeneous system over F of m equations and n unknowns. Let $\mathcal{W}$ be the set of all solutions in $\mathcal{F}^n$. Then $\mathcal{W}$ is a subspace of $\mathcal{F}^n$; that is, $\mathcal{W} \leqq \mathcal{F}^n$.

Proof. We know that $\mathcal{W}$ is non-empty; $\vec{0}$ is a valid solution to the homogeneous system, as $\mathbf{A}\vec{0} = \vec{0}$.

We can also show that $\mathcal{W}$ is closed under linear combinations; let $\vec{x}, \vec{y} \in \mathcal{W}$ and $c_1, c_2 \in F$; we have

$$\mathbf{A}(c_1\vec{x} + c_2\vec{y}) = c_1\mathbf{A}\vec{x} + c_2\mathbf{A}\vec{y} = c_1\vec{0} + c_2\vec{0} = \vec{0}.$$

$\square$

Now we can ask linear algebra questions about $\mathcal{W}$.

What is $\dim \mathcal{W}$? We can create a linear transformation associated to $\mathbf{A}$. The linear transformation T that we want to create must map some vectors in $\mathcal{F}^n$ to the zero vector in $\mathcal{F}^m$.

We let $T : \mathcal{F}^n \to \mathcal{F}^m$ where $\vec{v} \mapsto \mathbf{A}[\vec{v}]_e$, and $[T]_{e_{\mathcal{F}^n}}^{e_{\mathcal{F}^m}} = \mathbf{A}_{m\times n}$.

We have that T is linear and $\operatorname{Ker} T = \{\vec{x} \in \mathcal{F}^n : \mathbf{A}\vec{x} = \vec{0}\} = \mathcal{W}$.

This means that the solution space $\mathcal{W}$ to $\mathbf{A}\vec{x} = \vec{0}$ is precisely $\operatorname{Ker} T = \mathcal{W}$. This means that $\dim \mathcal{W} = \dim \operatorname{Ker} T = \operatorname{null} T$. Recall that the dimension theorem for T gives us that $\dim \mathcal{W} = n - \operatorname{rank} T$.

Lemma 15.5

The solution space to $\mathbf{A}\vec{x} = \vec{0}$ is a vector subspace of $\mathcal{F}^n$ of dimension $n - \operatorname{rank} T$.

Recall that $[T]_{e_{\mathcal{F}^n}}^{e_{\mathcal{F}^m}} = \mathbf{A}$. To find $\operatorname{rank}\mathbf{A}$, we can see that

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m_1} & a_{m_2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ T(\vec{e}_1) & T(\vec{e}_2) & \cdots & T(\vec{e}_n) \\ | & | & & | \end{bmatrix}.$$

Definition 15.6: Rank

$\operatorname{rank}\mathbf{A}$ is the maximum number of linearly independent $\vec{v}_i$'s, which we know that $\dim\operatorname{span}\{\vec{v}_1, \ldots, \vec{v}_n\} = \dim\operatorname{span}\{\text{columns of } \mathbf{A}\} = \dim \Im T = \operatorname{rank} T$.

Lemma 15.7

If $\mathbf{A}, \mathbf{B} \in \mathcal{M}_{m\times n}(F)$ are two matrices of the same linear transformation $T : \mathcal{F}^n \to \mathcal{F}^m$ (in different bases) then $\operatorname{rank}\mathbf{A} = \operatorname{rank}\mathbf{B} = \operatorname{rank} T = \dim \Im T$.

Moreover, $\operatorname{rank}\mathbf{A} \le m$ and $\operatorname{rank}\mathbf{A} \le n$.

Proof. We know that $\dim \Im T \le m$ because we're in the co-domain. But $\dim \Im T = \operatorname{rank} T$, so $\operatorname{rank} T \le m$.

Using the dimension theorem, we have that $\dim \Im T = n - \operatorname{null} T \le n$ as well. □

Let us go back to the homogeneous system $\mathbf{A}\vec{x} = \vec{0}$ where $\mathbf{A} \in \mathcal{M}_{m\times n}(F)$.

We know that the solution set $\mathcal{W}$ is a subspace of $\mathcal{F}^n$ with $\dim \mathcal{W} = n - \operatorname{rank} T = n - \operatorname{rank}\mathbf{A} \ge 0$. There are three cases:

- *Case 1*: $m < n$; fewer equations than variables

  In this case, we know that $\operatorname{rank}\mathbf{A} \le m < n$. This means that $\dim \mathcal{W} = n - \operatorname{rank}\mathbf{A} > 0$, and $\operatorname{Ker} T$ is nonzero.

  In this case, there will be infinitely many solutions (unless we work in a finite field); $\dim \mathcal{W} > 0$.

- *Case 2*: $m = n$; same number of equations and variables

  In one sub-case, we have exactly one solution, the zero vector $\vec{0}$. This will happen if and only if $\operatorname{Ker} T = \{\vec{0}\}$, and as such $\operatorname{rank}\mathbf{A} = n$ (the maximal rank). This can only happen if $\mathbf{A}$ is invertible, i.e. $\det\mathbf{A} \ne 0$.

  In another sub-case, we have more than one solution (that is, we have some non-trivial $\vec{x}$ such that $\mathbf{A}\vec{x} = \vec{0}$). This happens if and only if $\operatorname{rank}\mathbf{A} < n$ (the non-maximal rank). This can only happen if $\mathbf{A}$ is non-invertible, i.e. $\det\mathbf{A} = 0$; this allows $\operatorname{Ker}\mathbf{A}$ to include nonzero vectors.

- *Case 3*: $m > n$; next time.

15.2 Solving non-homogeneous systems

We want $\mathbf{A}\vec{x} = \vec{b} \ne \vec{0}$ where $\mathbf{A} \in \mathcal{M}_{m\times n}(F)$.

Let $\vec{x}_0$ be a particular solution to $\mathbf{A}\vec{x} = \vec{b}$; that is, $\mathbf{A}\vec{x}_0 = \vec{b}$ where $\vec{x}_0 \in \mathcal{F}^n$.

Let $\vec{x}_h$ be the general solution to the associated homogeneous system $\mathbf{A}\vec{x} = \vec{0}$.

What happens if we plug in $\vec{x}_0 + \vec{x}_h$? We have

$$\mathbf{A}(\vec{x}_0 + \vec{x}_h) = \mathbf{A}\vec{x}_0 + \mathbf{A}\vec{x}_h = \vec{b} + \vec{0} = \vec{b}.$$

This means that $\vec{x}_0 + \vec{x}_h$ is a solution to the non-homogeneous system. This means that we have constructed several solutions by fixing $\vec{x}_0$ and varying $\vec{x}_h$.

Conversely, let $\vec{y}$ be a solution to $\mathbf{A}\vec{x} = \vec{b}$; we want to show that $\vec{y} = \vec{x}_0 + \vec{x}_h$ for some $\vec{x}_h$. This is equivalent to showing that $\vec{y} - \vec{x}_0$ is a homogeneous solution. Plugging in, we have

$$\mathbf{A}(\vec{y} - \vec{x}_0) = \mathbf{A}\vec{y} - \mathbf{A}\vec{x}_0 = \vec{b} - \vec{b} = \vec{0}.$$

This means that $\vec{y} - \vec{x}_0$ is indeed a homogeneous solution $\vec{x}_h$, and $\vec{y} = \vec{x}_0 + \vec{x}_h$.

Theorem 15.8: Set of Solutions to a Non-Homogeneous System

The set of solutions to $\mathbf{A}\vec{x} = \vec{b}$ is the space of solutions to $\mathbf{A}\vec{x} = \vec{0}$ summed with $\vec{x}_0$, a particular solution to $\mathbf{A}\vec{x} = \vec{b}$.

Geometrically, we can get the set of solutions to $\mathbf{A}\vec{x} = \vec{b}$ by translating the space of homogeneous solutions by a particular non-homogeneous solution.

10/4/2021

Lecture 16

Systems of Linear Equations in Practice

When does a non-homogeneous system $\mathbf{A}\vec{x} = \vec{b}$ have a solution?

The non-homogeneous system has a solution iff $\vec{b} \in \operatorname{Im} T = \operatorname{span}\{\vec{v}_1, \dots, \vec{v}_n\}$. Notice that this is only true if we can add $\vec{b}$ and the span stays the same; that is, $\operatorname{span}\left\{\vec{v}_1, \dots, \vec{v}_n, \vec{b}\right\}$. This suggests that $\operatorname{rank}\mathbf{A} = \operatorname{rank}(\mathbf{A} \mid \vec{b})$

Theorem 16.1: Solutions to a Non-Homogeneous System

The non-homogeneous system has a solution if and only if $\operatorname{rank}\mathbf{A} = \operatorname{rank}(\mathbf{A} \mid \vec{b})$. We will call the RHS an augmented matrix of $\mathbf{A}\vec{x} = \vec{b}$.

How "many" solutions does $\mathbf{A}\vec{x} = \vec{b}$ have? That is, what is the dimension of the solution set? Or, when is there exactly one solution?

Recall Theorem 15.8; we need a particular solution to the non-homogeneous system. This is what we'll focus on today.

There is at least one solution if $\operatorname{rank}\mathbf{A} = \operatorname{rank}(\mathbf{A} \mid \vec{b})$. The "number" of solutions can be described by the dimension of the solution space.

We can add on to the table from last time for homogeneous systems (we let T be the linear transformation $T : \mathcal{F}^n \to \mathcal{F}^m$).

- *Case 1*: $m = n$.

  There is exactly one solution ($\vec{0}$) if and only if:

  - T is one-to-one; $\operatorname{Ker} T = \{\vec{0}\}$
  - T is onto; $\Im T = \mathcal{F}^n$
  - $\operatorname{rank} T = \operatorname{rank}\mathbf{A} = n$ (maximal)
  - T (or $\mathbf{A}$) is invertible
  - T is an isomorphism

There is more than one solution otherwise; when T is not invertible, or if $\operatorname{Ker} T \neq \{\vec{\mathbf{0}}\}$.

- *Case 2*: $m < n$.

  We know that $T = n - \operatorname{rank} T$. We further know that $\operatorname{rank} T < n$ because we're mapping to a smaller space. This means that $T > 0$, so $\operatorname{Ker} T$ is not trivial.

  This means that we always have more than one solution.

- *Case 3*: $m > n$.

  There is exactly one solution ($\vec{\mathbf{0}}$) if and only if $\operatorname{Ker} T = \{\vec{\mathbf{0}}\}$, and $\operatorname{rank}\mathbf{A} = n$ (maximized).

  There is more than one solution if and only if $\operatorname{rank}\mathbf{A} < n$; this follows similar reasoning from the previous case.

Example 16.2

$$\begin{cases} x_1 - 4x_2 - x_3 + x_4 = 3 \\ 2x_1 - 8x_2 + x_3 - 4x_4 = 9 \\ -x_1 + 4x_2 - 2x_3 + 5x_4 = -6 \end{cases}.$$

We want $\mathbf{A}\vec{\boldsymbol{x}} = \vec{\boldsymbol{b}} = \begin{bmatrix} 3 \\ 9 \\ -6 \end{bmatrix}$. Augmenting the coefficient matrix, we have

$$\mathbf{A} \mid \vec{\boldsymbol{b}} = \left[\begin{array}{cccc|c} 1 & -4 & -1 & 1 & 3 \\ 2 & -8 & 1 & -4 & 9 \\ -1 & 4 & -2 & 5 & -6 \end{array}\right].$$

We expect more than one solution because the homogeneous system always has more than one solution—however, $\vec{\boldsymbol{b}}$ may not be in the image, and we could have no solutions.

What is $\operatorname{rank}\mathbf{A}$? We want to apply column operations to $\mathbf{A}$ (we haven't shown whether column operations give the same result). We can do column operations here:

$$\left[\begin{array}{cccc|c} 1 & -4 & -1 & 1 & 3 \\ 2 & -8 & 1 & -4 & 9 \\ -1 & 4 & -2 & 5 & -6 \end{array}\right] \Longrightarrow \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 2 & 0 & 3 & -6 & 3 \\ -1 & 0 & -3 & 6 & -2 \end{array}\right] \qquad \begin{aligned} C_2 &\leftarrow C_2 + 4C_1 \\ C_3 &\leftarrow C_3 + C_1 \\ C_4 &\leftarrow C_4 - C_1 \\ C_5 &\leftarrow \frac{1}{3}C_5 \end{aligned}$$

$$\Longrightarrow \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 2 & 0 & 1 & 0 & 3 \\ -1 & 0 & -1 & 0 & -2 \end{array}\right]$$

We know that the columns that are left are linearly independent, and the last column is a linear combination of these remaining vectors. Since we got these resulting vectors from row operations (linear combinations), $\vec{\boldsymbol{b}}$ must be in the image.

We know that $\operatorname{rank}\mathbf{A} = 2$, so $\dim \mathcal{W} = 4 - 2 = 2$, so the solution space has dimension 2.

We also know that $\operatorname{rank}(\mathbf{A} \mid \vec{\boldsymbol{b}}) = 2 = \operatorname{rank}\mathbf{A}$, so $\mathbf{A}\vec{\boldsymbol{x}} = \vec{\boldsymbol{b}}$ has infinitely many solutions.

16.1 Column Operations

Which column operations can we use?

1. $C_i \mapsto aC_i$ $(a \neq 0)$

2. $C_i \mapsto C_i + aC_j$
3. $C_i \leftrightarrow C_j$

Why are these valid? They preserve the span of the columns of $\mathbf{A}$. Why do we care about preserving the span? These operations would then preserve $\operatorname{rank}\mathbf{A}$.

16.2 Solving a Homogeneous System

We apply row operations to $\mathbf{A}$ to get an upper triangular matrix or we can go to the RREF.

Example 16.3

$$\left[\begin{array}{cccc|c} 1 & -4 & -1 & 1 & 3 \\ 2 & -8 & 1 & -4 & 9 \\ -1 & 4 & -2 & 5 & -6 \end{array}\right] \implies \left[\begin{array}{cccc|c} 1 & -4 & -1 & 1 & 3 \\ 0 & 0 & 3 & -6 & 3 \\ 0 & 0 & -3 & 6 & -3 \end{array}\right] \qquad \begin{array}{l} R_1 \leftarrow R_1 - 2R_1 \\ R_3 \leftarrow R_3 + R_1 \end{array}$$

$$\implies \left[\begin{array}{cccc|c} 1 & -4 & -1 & 1 & 3 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \qquad \begin{array}{l} R_3 \leftarrow R_3 + R_2 \\ R_2 \leftarrow \frac{1}{3} R_2 \end{array}$$

This is currently in upper triangular form, but not yet RREF; we can still solve this by using back-substitution.

$$\implies \left[\begin{array}{cccc|c} 1 & -4 & 0 & -1 & 4 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \qquad R_1 \leftarrow R_1 + R_2$$

Now, this is in RREF form; $\mathbf{B} = \operatorname{RREF}(\mathbf{A})$. We can now solve $\mathbf{B}\vec{\boldsymbol{x}} = \vec{\mathbf{0}}$ instead.

We can conclude that $\mathbf{A}\vec{\boldsymbol{x}} = \vec{\mathbf{0}}$ has exactly the same solutions as $\mathbf{B}\vec{\boldsymbol{x}} = \vec{\mathbf{0}}$. We further have that rank is preserved:

$$\operatorname{rank}\mathbf{A} = n - \dim(\text{hom. solutions}) = \operatorname{rank}\mathbf{B}.$$

This is because we never changed the set of homogeneous solutions. Hence, row operations do not change $\operatorname{rank}\mathbf{A}$.

Example 16.4

Going back to the previous system, let us actually solve $\mathbf{B}\vec{\boldsymbol{x}} = \vec{\mathbf{0}}$:

$$\begin{cases} x_1 - 4x_2 \qquad - \; x_4 = 0 \\ \qquad\qquad x_3 - 2x_4 = 0 \end{cases} \implies \begin{cases} x_1 = 4x_2 + \; x_4 \\ x_3 = \qquad 2x_4 \end{cases}.$$

We have that $\{x_1, x_3\}$ are the leftmost variables in the rows, the single variables in the columns, and also called the leading variables.

We also have that $\{x_2, x_4\}$ are the remaining variables; they are non-leading variables and are called parameters/free variables.

If we let $x_2 = t_1$, $x_4 = t_2 \in F$, we have

$$\vec{\boldsymbol{x}} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4t_1 + t_2 \\ t_1 \\ 2t_2 \\ t_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ 0 \\ 0 \end{bmatrix} t_1 + \begin{bmatrix} 1 \\ 0 \\ 2 \\ 1 \end{bmatrix} t_2.$$

These two vectors are a basis for the solution space of $\mathbf{A}\vec{\boldsymbol{x}} = \vec{\mathbf{0}}$ in $\mathcal{F}^n$.

How do we solve the system $\mathbf{B}\vec{x} = \vec{c}$? We have

$$\begin{cases} x_1 - 4x_2 \qquad - \ x_4 = 0 \\ \qquad\qquad x_3 - 2x_4 = 0 \end{cases} \implies \vec{x}_0 = \begin{bmatrix} 4 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

We can set the free variables to zero, and we can read off the particular non-homogeneous solution.

This means that all solutions to $\mathbf{A}\vec{x} = \vec{b}$ are of the form

$$\vec{x} = \begin{bmatrix} 4 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 4 \\ 1 \\ 0 \\ 0 \end{bmatrix} t_1 + \begin{bmatrix} 1 \\ 0 \\ 2 \\ 1 \end{bmatrix} t_2, \qquad t_1, t_2 \in F.$$

16.3 RREF

What are the properties of an RREF?

The first nonzero entry in each row is 1: this is called the leading 1. Further, there are no other nonzero entries in the columns of a leading 1. We further have a NW/SE staircase of 1's, and all zero rows are at the bottom.

Let us compare rank RREF and rank $\mathbf{A}$; we know they are equal because this does not change the solution space.

We have

$$\begin{aligned} \operatorname{rank}\mathbf{A} &= \#\text{ LI columns in } \mathbf{A} \\ &= \#\text{ leading columns in RREF} \\ &= \#\text{ leading 1's in RREF} \\ &= \#\text{ leading rows in RREF} \\ &= \#\text{ linearly independent rows in RREF} \\ &= \#\text{ LI rows in } \mathbf{A} \end{aligned}$$

This means that $\operatorname{rank}\mathbf{A} = \#$ LI columns in $\mathbf{A} = \#$ LI rows in $\mathbf{A}$. Hence, $\operatorname{rank}\mathbf{A} = \operatorname{rank}\mathbf{A}^T$.

Is the RREF($\mathbf{A}$) unique? Yes, but the proofs are nontrivial.

10/6/2021

Lecture 17

Determinants

17.1 Determinants

Determinants arise naturally when solving linear systems of equations.

Example 17.1

$$\begin{vmatrix} \sin\alpha & \cos\alpha \\ -\cos\alpha & \sin\alpha \end{vmatrix} = \sin^2\alpha + \cos^2\alpha = 1$$

$$\begin{vmatrix} x-1 & 1 \\ x^3 & x^{2+x+1} \end{vmatrix} = (x-1)(x^2+x+1) - x^3.$$

Using the fact that $x^3 - y^3 = (x-y)(x^2+xy+y^2)$, we have that this determinant is just $x^3 - 1 - x^3 = -1$.

The determinants (denoted as Δ) in the previous examples are called determinants of order 2.

A determinant of order three looks like

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}.$$

Notice that these are all possible products of numbers, no two in the same row or column.

We can count these as 3! (3 choices for row 1, 2 choices for row 2, and 1 choice for row 3), or in general $n!$ for a determinant of order n.

Calculations are omitted for brevity, but we can use determinants to solve systems of equations with Cramer's rule:

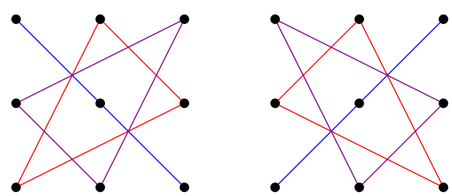
Theorem 17.2: Cramer's Rule

If $\Delta \neq 0$, then $\Delta \cdot x_i = \Delta_j$:

$$x_1 = \frac{\Delta_1}{\Delta}, \quad x_2 = \frac{\Delta_2}{\Delta}, \quad x_3 = \frac{\Delta_3}{\Delta}.$$

Here, Δ_i denotes the coefficient matrix with the ith column replaced by the constant vector $\vec{\boldsymbol{b}}$.

17.1.1 Triangle Rule



We have 3 products that are positive (with one side parallel to the main diagonal), and 3 products that are negative (with one side parallel to the anti-diagonal).

17.2 Determinants of Order n

The determinant of a $n \times n$ matrix $\mathbf{A}$ is the sum of all possible products of entries in a non-attacking rook configuration, some with =, and some with −.

The direct formula for the determinant is:

$$\sum_{\alpha_1, \alpha_2, \ldots, \alpha_n} \pm a_{1\alpha_1} a_{2\alpha_2} \cdots a_{n\alpha_n}.$$

Here, we're summing over all possible permutations of $(1, 2, \ldots, n)$.

All that is left is the signs—which terms get + and which terms get −?

Definition 17.3: Inversions

If $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ is a permutation of $(1, 2, \ldots, n)$, then an inversion is a pair (α_i, α_j) that is out of order (that is, $\alpha_i > \alpha_j, i < j$).

For example, in the case of $n = 3$, we have

$n = 3$	inversions	parity	sign
(1, 2, 3)	0	even	+
(2, 1, 3)	1	odd	−
(2, 3, 1)	2	even	+
(1, 3, 2)	1	odd	−
(3, 1, 2)	2	even	+
(3, 2, 1)	3	odd	−

The permutations with an even parity are positive, and the permutations with an odd parity are negative; that is, the sign is $(-1)^k$, where k is the number of inversions in the permutation.

Notation-wise, we will say that $[\alpha_1, \alpha_2, \dots, \alpha_n]$ denotes the number of inversions in $(\alpha_1, \alpha_2, \dots, \alpha_n)$.

Definition 17.4: Sign of Determinant Term

The sign in front of $a_{1\alpha_1} a_{2\alpha_2} \cdots a_{n\alpha_n}$ in Δ corresponds to $(-1)^{[\alpha_1,\dots,\alpha_n]}$.

Definition 17.5: General Definition of Determinant

Δ of order n is equal to

$$\Delta = \sum_{\text{perm. } (\alpha_1,\dots,\alpha_n)} (-1)^{[\alpha_1,\dots,\alpha_n]} a_{1\alpha_1} a_{2\alpha_2} \cdots a_{n\alpha_n}.$$

We can recognize that the determinant Δ of order n is an polynomial of degree n with $n!$ terms, half with + and half with −. Each term corresponds to a non-attacking rook configuration.

Example 17.6

$$\begin{vmatrix} 0 & 2 & 0 & 0 \\ 1 & 7 & 4 & 5 \\ 0 & 11 & 3 & 0 \\ 0 & 5 & 0 & 1 \end{vmatrix} = (-1) \cdot 2 \cdot 1 \cdot 3 \cdot 1 = -6.$$

Notice that this is the only nonzero term in the determinant expression; other terms will contain a zero.

One tip in recognizing inversions is that lines with positive slope are inversions, and lines with negative slope are non-inversions.

10/8/2021

Lecture 18

Properties of Determinants

review lecture video

10/11/2021

Lecture 19

Introduction to Eigenvectors and Eigenvalues

Recall the reflection problem we looked at in Example 13.7; we were essentially looking at $T(\vec{v}) = \lambda \vec{v}$ for some $\lambda \in \mathbb{R}$. In the example, $\lambda_1 = 1$, $\lambda_2 = -1$.

Definition 19.1: Eigenvectors and Eigenvalues

Let $T : \mathcal{V}_F \to \mathcal{V}_F$ be a linear operator. Then a vector $\vec{\boldsymbol{v}} \in \mathcal{V}$ where $\vec{\boldsymbol{v}} \neq \vec{\boldsymbol{0}}$, such that $T(\vec{\boldsymbol{v}}) = \lambda \vec{\boldsymbol{v}}$ for some $\lambda \in F$ is called an *eigenvector* of T corresponding to eigenvalue λ.

Example 19.2

Suppose we have $T : \mathbb{R}^2 \to \mathbb{R}^2$, $[T] = \begin{bmatrix} -1 & 2 \\ -3 & 4 \end{bmatrix}$.

Suppose we have our "divine" guess of $\vec{\boldsymbol{v}}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\vec{\boldsymbol{v}}_2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$.

To check whether these are eigenvectors, we have

$$T(\vec{\boldsymbol{v}}_1) = \begin{bmatrix} -1 & 2 \\ -3 & 4 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1 \cdot \vec{\boldsymbol{v}}_1$$

$$T(\vec{\boldsymbol{v}}_2) = \begin{bmatrix} -1 & 2 \\ -3 & 4 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix} = 2\vec{\boldsymbol{v}}_2$$

Further, we have a basis $\beta = \{\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2\}$, and $[T]_\beta = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}_\beta$.

Theorem 19.3: Diagonalizability with Eigenbasis

Let $T : \mathcal{V} \to \mathcal{V}$ be a linear operator. Then T has a diagonal matrix $\mathbf{A} = [T]_\beta$ in some basis β if and only if β consists entirely of eigenvectors (and β is called an *eigenbasis* for T).

Proof. ($\Longrightarrow$) Let T have a diagonal matrix $[T]_\beta = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_2 \end{bmatrix}$, where $\beta = \{\vec{\boldsymbol{v}}_1, \dots, \vec{\boldsymbol{v}}_n\}$ is a basis.

We have

$$T(\vec{\boldsymbol{v}}_i) = \begin{bmatrix} \lambda_1 & \cdots & \cdots & 0 & 0 \\ \vdots & \ddots & & \ddots & 0 \\ \vdots & & \lambda_i & & \vdots \\ 0 & \ddots & & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & \lambda_n \end{bmatrix}\begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ \lambda_i \\ \vdots \\ 0 \end{bmatrix} = \lambda_i \vec{\boldsymbol{v}}_i.$$

($\Longleftarrow$) Suppose $\beta = \{\vec{\boldsymbol{v}}_1, \dots, \vec{\boldsymbol{v}}_n\}$ is an eigenbasis. This means that we have

$$[T]_\beta = \begin{bmatrix} | & & | \\ T(\vec{\boldsymbol{v}}_1)_\beta & \cdots & T(\vec{\boldsymbol{v}}_n)_\beta \\ | & & | \end{bmatrix}$$

$$T(\vec{\boldsymbol{v}}_i) = \lambda_i \vec{\boldsymbol{v}}_i = \lambda_i \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}_\beta$$

This gives us our diagonal matrix of λ's. $\square$

Example 19.4

What is the eigenbasis for $\mathcal{I}_{\mathcal{V}}$? Any basis is an eigenbasis, as $\mathcal{I}_{\mathcal{V}}(\vec{\boldsymbol{v}}) = 1 \cdot \vec{\boldsymbol{v}}$. Further, $[\mathcal{I}_{\mathcal{V}}]_\beta = \mathbf{I}_n$, as we must have 1's along the diagonal.

What is the eigenbasis for $0_{\mathcal{V}}$? Again any vector is an eigenbasis; $0_{\mathcal{V}}(\vec{\boldsymbol{v}}) = 0 \cdot \vec{\boldsymbol{v}}$.

How do we find an eigenbasis for T? We change gears; we search for eigenvalues first!

Example 19.5

From before, we're looking for $\vec{\boldsymbol{v}} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \neq \vec{\mathbf{0}}$ such that $T(\vec{\boldsymbol{v}}) = \lambda \vec{\boldsymbol{v}}$ for some $\lambda \in \mathbb{R}$.

$$\begin{bmatrix} -1 & 2 \\ -3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \mathcal{I}_{\mathcal{V}} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Notice that we can factor stuff out and simplify:

$$\left(\begin{bmatrix} -1 & 2 \\ -3 & 4 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$\underbrace{\begin{bmatrix} -1-\lambda & 2 \\ -3 & 4-\lambda \end{bmatrix}}_{\mathbf{B}} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

This is a 2×2 homogeneous system with a solution $\vec{\boldsymbol{x}} \neq 0$ such that $\mathbf{B}\vec{\boldsymbol{x}} = \vec{\mathbf{0}}$.

This system has a non-trivial solution if and only if $\mathbf{B}$ is non-invertible, which is true if and only if $\det \mathbf{B} = 0$.

In our case, we have

$$\det \mathbf{B} = (-1-\lambda)(4-\lambda) + 6 = \lambda^2 - 3\lambda + 2 = (\lambda - 1)(\lambda - 2) = 0.$$

This means that $\lambda_1 = 1$, $\lambda_2 = 2$ are the only possible eigenvalues for T.

Now, we have to find the eigenvectors; we have two cases:

- *Case 1*: $\lambda_1 = 1$, Solving $\mathbf{B}\vec{\boldsymbol{x}} = \vec{\mathbf{0}}$. This is a modified homogeneous system;

$$\left[\begin{array}{cc|c} -1-1 & 2 & 0 \\ -3 & 4-1 & 0 \end{array}\right] \implies \left[\begin{array}{cc|c} -2 & 2 & 0 \\ -3 & 3 & 0 \end{array}\right]$$

  This gives us $\vec{\boldsymbol{v}} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, for example $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ because $x_1 = x_2$.

- *Case 2*: $\lambda_2 = 2$.

  Solving the system (omitted here), we have $\vec{\boldsymbol{v}} = \begin{bmatrix} x_1 \\ \frac{3}{2}x_1 \end{bmatrix}$, for example $\vec{\boldsymbol{v}}_2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$.

  As such, if $\vec{\boldsymbol{v}}_1 = \begin{bmatrix} a \\ a \end{bmatrix}$ and $\vec{\boldsymbol{v}}_2 = \begin{bmatrix} 2b \\ 3b \end{bmatrix}$, for some $a, b \neq 0$, then $\beta = \{\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2\}$ is an eigenbasis for T corresponding to eigenvalues 1 and 2, with $[T]_\beta = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$.

19.1 General Algorithm to find Eigenvalues

Suppose we have $\mathbf{A}_{n\times n} = [T]_\gamma$. We first let $\mathbf{B}_{n \times n} = \mathbf{A} - \lambda \mathbf{I}_n$.

The eigenvalues for T will be precisely all roots of $\det \mathbf{B}$, which is a polynomial in terms of λ, where we want to solve when equal to zero.

This will expand out to $(-1)^n\lambda^n +$ lower degree terms in λ.

Definition 19.6: Characteristic Polynomial

Let $\mathbf{A} \in \mathcal{M}_{n\times n}(F)$. The *characteristic polynomial* of $\mathbf{A}$ is defined as

$$p(\lambda) = \text{char}_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I}).$$

This is a polynomial of degree n with leading coefficient $(-1)^n$.

Theorem 19.7: Computing Eigenvalues and Eigenvectors

Let $T : \mathcal{V} \to \mathcal{V}$ be a linear operator, $\mathbf{A}$ be the matrix of T in some basis. Then

- The eigenvalues of T are the roots of $\text{char}_{\mathbf{A}}(\lambda)$
- For an eigenvalue λ of T, all eigenvectors $\vec{\boldsymbol{v}}$ corresponding to λ are the nonzero vectors in $\text{Ker}(\mathbf{A} - \lambda \mathbf{I})$;

$$(\mathbf{A} - \lambda \mathbf{I})\vec{\boldsymbol{v}} = \vec{\mathbf{0}} \implies \mathbf{A}\vec{\boldsymbol{v}} = \lambda \vec{\boldsymbol{v}}.$$

Example 19.8

Let $[T]_e = \begin{bmatrix} 0 & -2 & -3 \\ -1 & 1 & -1 \\ 2 & 2 & 5 \end{bmatrix}$. Find the eigenbasis for T.

IF we solve $\det(\mathbf{A} - \lambda \mathbf{I}) = \vec{\mathbf{0}}$, then we have $\lambda_1 = 1$, $\lambda_2 = 2$, and $\lambda_3 = 3$.

We could be clever and notice repeated factors as we expand, but here is one trick. The characteristic polynomial turns out to be $-\lambda^3 + 6\lambda^2 - 11\lambda + 5 = 0$. Notice that this only has integer coefficients; the first coefficient is $a_n = -1$ and the last is $a_0 = 6$.

Theorem 19.9: Rational Root Theroem

If $r = \frac{a}{b} \in \mathbb{Q}$ is a root, then $a \mid a_0$ and $b \mid a_n$.

Here, $a \mid 6$ and $b \mid -1$. WLOG, suppose $b = 1$. We have $a \in \{\pm 1, \pm 2, \pm 3, \pm 6\}$. We have eight possible rational roots $r = \pm 1, \pm 2, \pm 3, \pm 6$. We check by brute force, plugging them in.

Note that once we've found one, we can factor it out and repeat, decreasing the number of possibilities, and when we get down to a quadratic, we can solve directly.

If we went on and found the eigenvectors, we'd have

$$\vec{\boldsymbol{v}}_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \qquad \vec{\boldsymbol{v}}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \qquad \vec{\boldsymbol{v}}_3 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}.$$

(Calculations omitted for brevity.)

Lemma 19.10

The characteristic polynomial of $T : \mathcal{V} \to \mathcal{V}$ does not depend on the choice of the matrix $[T]_\beta = \mathbf{A}$. That is, $\text{char}_T(\lambda) = \text{char}_{\mathbf{A}}(\lambda) = \text{char}_{\mathbf{B}}(\lambda)$ for $[T]_\beta = \mathbf{A}$, $[T]_\gamma = \mathbf{B}$.

Proof. We have $\text{char}_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I})$, and notice that $\mathbf{A} \sim \mathbf{B}$, where $\mathbf{B} = \mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$, and $\mathbf{A} = \mathbf{Q}\mathbf{B}\mathbf{Q}^{-1}$.

We have

$$\begin{aligned}
\text{char}_{\mathbf{A}}(\lambda) &= \det\left(\mathbf{Q}\mathbf{B}\mathbf{Q}^{-1} - \lambda\mathbf{Q}\mathbf{I}\mathbf{Q}^{-1}\right) \\
&= \det\left(\mathbf{Q}\mathbf{B}\mathbf{Q}^{-1} - \lambda\mathbf{Q}\mathbf{I}\mathbf{Q}^{-1}\right) \\
&= \det\left(\mathbf{Q}(\mathbf{B} - \lambda\mathbf{I})\mathbf{Q}^{-1}\right) \\
&= \det(\mathbf{Q})\det(\mathbf{B} - \lambda\mathbf{I})\det\left(\mathbf{Q}^{-1}\right) \\
&= \det(\mathbf{B} - \lambda\mathbf{I}) \\
&= \text{char}_{\mathbf{B}}(\lambda)
\end{aligned}$$

Here, we used some clever multiplication of $\mathbf{Q}$ and $\mathbf{Q}^{-1}$ around the identity, and also used the fact that the determinant of a product is the product of the determinants, cancelling out the determinants of $\mathbf{Q}$ and $\mathbf{Q}^{-1}$. □

10/13/2021

Lecture 20

Diagonalization I: Geometric Multiplicities

Definition 20.1: Eigenspace

Let $T : \mathcal{V} \to \mathcal{V}$ be a linear operator and let α be an eigenvalue for T. Then the space of all eigenvectors of T corresponding to α is called the eigenspace E_α of T:

$$E_\alpha = \{\vec{\boldsymbol{v}} \in \mathcal{V} \mid T(\vec{\boldsymbol{v}}) = \alpha\vec{\boldsymbol{v}}\} = \text{Ker}(T - \alpha\mathcal{I}_{\mathcal{V}}).$$

Note that $\vec{\mathbf{0}} \in E_\alpha$ but $\vec{\mathbf{0}}$ is not an eigenvector.

We have that $T - \alpha\mathcal{I}_{\mathcal{V}}$ is a new linear operator, which we'll dub as the "Halloween" operator; it sends the corresponding eigenspace to $\vec{\mathbf{0}}$. For each eigenvalue, we have a different operator.

Example 20.2

If $T = \mathcal{I}_{\mathcal{V}} : \mathcal{V} \to \mathcal{V}$, then $\lambda = 1$ and, $E_1 = \mathcal{V}$.

If $T = 0_{\mathcal{V}} : \mathcal{V} \to \mathcal{V}$, then $\lambda = 0$, and $E_0 = \mathcal{V}$.

Example 20.3

If $T : \mathcal{V} \to \mathcal{V}$ such that $T^2 = T$ (idempotent), recall that $\mathcal{V} = \text{Ker}\, T \oplus \text{Im}\, T$, and T is the projection onto $\text{Im}\, T$ along $\text{Ker}\, T$.

What are E_λ for T? Let us find the eigenvalues of T.

We have $T(\vec{v}) = \lambda\vec{v}$, and since T is idempotent, we have

$$\begin{aligned} T^2(\vec{v}) &= T(\lambda\vec{v}) \\ T(\vec{v}) &= \lambda T(\vec{v}) && (\vec{v} \text{ eigenvector}) \\ \lambda\vec{v} &= \lambda^2\vec{v} \\ \lambda(\lambda-1)\vec{v} &= 0 && (\text{rearrange, factor}) \end{aligned}$$

Since $\vec{v} \neq 0$, we must have $\lambda = 0, 1$.

We thus have $E_0 = \{\vec{v} \in \mathcal{V} \mid T(\vec{v}) = \vec{0}\} = \operatorname{Ker} T$. Similarly, $E_1 = \{\vec{v} \in \mathcal{V} \mid T(\vec{v}) = \vec{v}\} = \operatorname{Im} T$.

This means that $\mathcal{V} = E_0 \oplus E_1$. Further, if we have a basis $\{\vec{v}_i\}$ for E_0, and $\{\vec{w}_j\}$ for E_1, then their union is a basis for $\mathcal{V}$ made up of eigenvectors.

This means that $[T]_\beta$ is a matrix with zeroes and ones along the diagonal; zeroes corresponding to bases of $\operatorname{Ker} T$, and ones corresponding to bases of $\operatorname{Im} T$.

Definition 20.4: Geometric Multiplicity

The *geometric multiplicity* of an eigenvalue λ of T is $m_g(\lambda) = \dim E_\lambda$.

Example 20.5

Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ with

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Computing $\mathbf{A} - \lambda\mathbf{I}$, we have

$$\mathbf{A} - \lambda\mathbf{I} = \begin{bmatrix} 1-\lambda & 1 & 0 \\ 0 & 1-\lambda & 0 \\ 0 & 0 & 3-\lambda \end{bmatrix} \implies \det(\mathbf{A} - \lambda\mathbf{I}) = (1-\lambda)(1-\lambda)(3-\lambda).$$

The latter fact is because the matrix is upper triangular. This means that $\lambda_1 = 1$ and $\lambda_2 = 3$

We have that $E_1 = \operatorname{Ker}(T - \mathcal{I}_\mathcal{V})$ gives us $x_3 = 0 = x_2$, meaning

$$E_1 = \{t \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \mid t \in \mathcal{R}\}.$$

With $E_3 = \operatorname{Ker}(T - 3\mathcal{I}_\mathcal{V})$, we have $x_2 = 0 = x_1$, meaning

$$E_2 = \{t \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \mid t \in \mathbb{R}\}.$$

This means that $m_g(1) = \dim E_1 = 1$, and $m_g(3) = \dim E_3 = 1$.

Since $m_g(1) + m_g(3) \neq \dim \mathbb{R}^3$, Our eigenbasis has only dimension 2, so we can't create an eigenbasis for $\mathbb{R}^3$. This means that T is not diagonalizable.

However, notice that $\vec{v}_1 \in E_1$ and $\vec{v}_2 \in E_3$ are linearly independent! This makes us investigate; what happens when eigenvectors correspond to distinct eigenvalues?

Theorem 20.6: Linear Independence of Eigenvectors for Distinct Eigenvalues

Let $T : \mathcal{V} \to \mathcal{V}$ be linear operator, with $\lambda_1, \ldots, \lambda_k$ be distinct eigenvalues for T, with $\vec{\boldsymbol{v}}_1, \ldots, \vec{\boldsymbol{v}}_k$ being their corresponding eigenvectors.

Then, $S = \{\vec{\boldsymbol{v}}_1, \ldots, \vec{\boldsymbol{v}}_k\}$ is linearly independent.

Proof. We proceed by induction on m, where k is fixed.

Base Case ($m = 1$) We look at $\{\vec{\boldsymbol{v}}_1\}$, which is linearly independent. It's an eigenvector, which is not zero.

Inductive Hypothesis: We assume that $\{\vec{\boldsymbol{v}}_1, \ldots, \vec{\boldsymbol{v}}_m\}$ are linearly independent for some $1 \le m < k$.

Inductive Step: We want to show that $\{\vec{\boldsymbol{v}}_1, \ldots, \vec{\boldsymbol{v}}_m, \vec{\boldsymbol{v}}_{m+1}\}$ is still linearly independent.

We have the linear relation

$$\begin{aligned} a_1\vec{\boldsymbol{v}}_1 + \cdots + a_m\vec{\boldsymbol{v}}_m + a_{m+1}\vec{\boldsymbol{v}}_{m+1} &= \vec{\boldsymbol{0}} \\ a_1 T(\vec{\boldsymbol{v}}_1) + \cdots + a_m T(\vec{\boldsymbol{v}}_m) + a_{m+1} T(\vec{\boldsymbol{v}}_{m+1}) &= \vec{\boldsymbol{0}} \\ a_1\lambda_1\vec{\boldsymbol{v}}_1 + \cdots + a_m\lambda_m\vec{\boldsymbol{v}}_m + a_{m+1}\lambda_{m+1}\vec{\boldsymbol{v}}_{m+1} &= \vec{\boldsymbol{0}} && (\vec{\boldsymbol{v}}_i \text{ eigenvectors}) \end{aligned}$$

If we multiply the first equation by λ_{m+1} and subtract the last equation from the first, we eliminate $\vec{\boldsymbol{v}}_{m+1}$ and we have

$$a_1(\lambda_{m+1} - \lambda_1)\vec{\boldsymbol{v}}_1 + \cdots + a_m(\lambda_{m+1} - \lambda_m)\vec{\boldsymbol{v}}_m = \vec{\boldsymbol{0}}$$

We have that $\vec{\boldsymbol{v}}_1$ through v_m are linearly independent, so all a_1 through a_m are all zero, because $\lambda_{m+1} - \lambda_i$ are not zero (they're distinct).

If all of the a_1 through a_m are zero, then we must have a_{m+1} is also zero, because $a_{m+1}\vec{\boldsymbol{v}}_{m+1} = \vec{\boldsymbol{0}}$ and $\vec{\boldsymbol{v}}_{m+1}$ cannot be $\vec{\boldsymbol{0}}$.

As such, $\{\vec{\boldsymbol{v}}_1, \ldots, \vec{\boldsymbol{v}}_{m+1}\}$ are linearly independent. □

Lemma 20.7

Let $T : \mathcal{V} \to \mathcal{V}$ be a linear operator, with $\dim \mathcal{V} = n$. If T has n distinct eigenvalues, then T is diagonalizable.

This is a sufficient but not necessary condition for diagonalizability.

Proof. Let $\{\vec{\boldsymbol{v}}_1, \ldots, \vec{\boldsymbol{v}}_n\}$ be eigenvectors corresponding to $\{\lambda_1, \ldots, \lambda_n\}$ (distinct).

This means that $\beta = \{\vec{\boldsymbol{v}}_1, \ldots, \vec{\boldsymbol{v}}_n\}$ are linearly independent, of the right dimension, and thus is an eigenbasis for $\mathcal{V}$. This means that T is diagonalizable:

$$[T]_\beta = \begin{bmatrix} \lambda_1 & & \cdots & 0 \\ \vdots & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & & \lambda_n \end{bmatrix}.$$

□

Note that not every diagonalizable operator has n distinct eigenvalues!

Example 20.8

$T = \mathcal{I}_\mathcal{V}$ has only 1 eigenvalue $\lambda_1 = 1$, but $[T]_\beta = \mathbf{I}_n$ in any basis. This means that T is diagonalizable (moreover, it is diagonal!); this means that the previous condition is not necessary.

Example 20.9

Let $T: \mathcal{M}_{n\times n}(\mathbb{R}) \to \mathcal{M}_{n\times n}(\mathbb{R})$, such that $T(\mathbf{A}) = \mathbf{A}^T$.

Is T diagonalizable? What if $U^2 = \mathcal{I}_\mathcal{V}$?

We know that $T^2(\mathbf{A}) = (\mathbf{A}^T)^T = \mathbf{A}$, and as such $T^2 = \mathcal{I}_\mathcal{V}$. Suppose λ is an eigenvalue and $\vec{\boldsymbol{v}}$ is the corresponding eigenvector.

We have $T^2(\vec{\boldsymbol{v}}) = \lambda^2\vec{\boldsymbol{v}}$, but also $T^2(\vec{\boldsymbol{v}}) = \mathcal{I}_\mathcal{V}(\vec{\boldsymbol{v}}) = \vec{\boldsymbol{v}}$.

This means that $(\lambda^2 - 1)\vec{\boldsymbol{v}}$, and $\lambda = \pm 1$.

We have two eigenspaces:

- E_1: $T(\mathbf{A}) = \mathbf{A}^T = 1\cdot\mathbf{A}$, and as such $\mathbf{A}$ must be symetric.

  This means that $E_1 = S_n$, the space of symmetric $n\times n$ matrices. Further, $\dim S_n = \frac{n(n+1)}{2}$.
- E_{-1}: $T(\mathbf{A}) = \mathbf{A}^T = -1\cdot\mathbf{A}$, and as such $\mathbf{A}$ must be skew-symmetric.

  This means that $E_{-1} = A_n$, the space of skew-symmetric $n\times n$ matrices. Further, $\dim A_n = \frac{n(n-1)}{2}$.

Further, we have that $\dim S_n + \dim A_n = n^2 = \dim \mathcal{M}_{n\times n}$.

Notice that $T(\mathbf{B}_1 + \mathbf{B}_2) = \mathbf{B}_1 - \mathbf{B}_2$ where $\mathbf{B}_1 \in S_n$ and $\mathbf{B}_2 \in A_n$. This means that T can be thought of as a reflection!

Some conclusions:

- T is diagonalizable
- The eigenbasis β consists of a basis for S_n and a basis for A_n.
- $[T]_\beta$ is a diagonal matrix with only 1's and -1's on the diagonal.

10/15/2021

Lecture 21

Diagonalization II: Algebraic Multiplicities

Today, we reverse the tables; if T is diagonalizable, then what must have happened?

Example 21.1

What happens to the identity transformation $\mathcal{I}_\mathcal{V}: \mathcal{V} \to \mathcal{V}$?

We have $\text{char}_T(\lambda) = \det(\mathbf{I} - \lambda\mathbf{I}) = (1-\lambda)^n$. The root $\lambda = 1$ appears n times! We'll call this the algebraic multiplicity.

In this situation, $m_g(1) = n = \dim\mathcal{V} = \dim E_1$ (because the eigenspace of $\lambda = 1$ is $\mathcal{V}$).

Here, we have that the algebraic multiplicity is the same as the geometric multiplicity of 1.

Definition 21.2: Algebraic Multiplicity in Polynomials

If $f(x)$ is a polynomial of degree n over a field F, then a root $x = a$ of f has multiplicity m if $f(x) = (x-a)^m \cdot g(x)$, where $g(a) \neq 0$. That is, a is not a root of g.

Definition 21.3: Algebraic Multiplicity of Eigenvalues

If $T : \mathcal{V} \to \mathcal{V}$ is a linear operator, with λ as an eigenvalue, then the algebraic multiplicity of λ is its multiplicity as a root of $\text{char}_T(x)$.

In particular, $\text{char}_T(x) = \det(\mathbf{A} - x\mathbf{I}) = (x-\lambda)^k \cdot g(x)$, and the algebraic multiplicity, denoted as $m_a(\lambda)$, is equal to k, where $g(\lambda) \neq 0$.

Example 21.4

Suppose $\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$.

We have $\text{char}_{\mathbf{A}}(x) = (x-1)^2 \cdot (3-x)^1$. We have $m_a(1) = 2$, and $m_g(1) = 1$ from earlier. Similarly, $m_a(3) = 1$ and $m_g(3) = 1$ from earlier.

Hence, we find that the algebraic and geometric multiplicities may not always be equal.

What do m_a and m_g Hae to do with diagonalization?

Suppose T is diagonalizable. This means that we have $\beta = \{\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2, \dots, \vec{\boldsymbol{v}}_n\}$ as an eigenbasis, such that $[T]_\beta$ is diagonal with the eigenvalues.

We then know that $\text{char}_T(x) = (x-\lambda_1)\cdots(x-\lambda_n)$, and as such it has n roots (but not necessarily distinct).

Definition 21.5: Splitting Polynomials

If a polynomial $f(x)$ completely factors into linear factors in F, we say that $f(x)$ splits over F.

Example 21.6

We can see that $f(x) = x^2 + 1$ does not split in $\mathbb{R}$, but it does split over $\mathbb{C}$. That is, we have $x^2 + 1 = (x-i)(x+i)$.

Similarly, suppose we have $[T]_e = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, where $\text{char}_T(x) = x^2 + 1$.

This means that T is not diagonalizable over $\mathbb{R}$, but T is diagonalizable over $\mathbb{C}$ because $\lambda_{1,2} = \pm i$, where the diagonal matrix is $\begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$. (Splits and has distinct eigenvalues $\implies$ diagonalizable)

Lemma 21.7

If $T : \mathcal{V}_F \to \mathcal{V}_F$ is a linear operator and is diagonalizable, then $\text{char}_T(x)$ splits over F.

This is a necessary but not sufficient criterion for diagonalizability.

Example 21.8

From before, taking $\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$, we know that $\text{char}_{\mathbf{A}}(x) = (x-1)^2 \cdot (3-x)^1$ splits over $\mathbb{R}$.

However, $\mathbf{A}$ is *not* diagonalizable!

What went wrong? If we compare m_a and m_g for each eigenvalue, we have again $m_a(1) = 2$ but $m_g(3) = 1$, and $m_a(3) = 1$ and $m_g(3) = 1$.

We can see that the multiplicities for 3 are equal, but the algebraic multiplicity is greater than the geometric multiplicity for 1.

Theorem 21.9: Criteria for Diagonalizability

The linear operator $T : \mathcal{V}_F \to \mathcal{F}_F$ is diagonalizable if and only if

- $\text{char}_T(x)$ splits over F
- $m_a(\lambda) = m_g(\lambda)$ for all eigenvalues λ of T.

This tells us that T has sufficiently many eigenvectors for a basis.

Example 21.10

Let $\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$. We have $\text{char}_{\mathbf{A}}(x) = (x-1)^1 \cdot (x+1)^1 \cdot (x-2)^2$.

We can already see that we split over $\mathbb{R}$. Looking at each eigenvalue, we have

- $m_a(1) = 1 = m_g(1)$ (because the geometric multiplicity must be at least 1)
- $m_a(-1) = 1 = m_g(1)$
- $m_a(2) = 2 = m_g(2)$ (because the bottom right 2×2 will become $\mathbf{0}$)

As such, we conclude that $\mathbf{A}$ is diagonalizable over $\mathbb{R}$, with the eigenvalues (along with their multiplicities) on the diagonal.

21.1 Criteria for Diagonalizability

We have that

- $T : \mathcal{V}_F \to \mathcal{V}_F$ diagonalizable $\implies \text{char}_T(x)$ splits over F
- $\text{char}_T(x)$ splits over $F \implies \sum_{i=1}^k m_a(\lambda_i) = n$
- $\lambda_1, \ldots, \lambda_k$ distinct $\implies \sum_{i=1}^k m_g(\lambda_i$
- $\sum_{i=1}^k m_g(\lambda_i) = n \implies m_a(\lambda_i) = m_g(\lambda_i)$
- $m - a(\lambda_i) = m_g(\lambda_i)$ and $\text{char}_T(x)$ splits over $F \implies T$ is diagonalizable.

We know that if $\vec{\boldsymbol{v}}_1 \in E_{\lambda_1}, \vec{\boldsymbol{v}}_2 \in E_{\lambda_2}, \ldots, \vec{\boldsymbol{v}}_k \in E_{\lambda_k}$, then $\vec{\boldsymbol{v}}_1, \ldots, \vec{\boldsymbol{v}}_k$ are linearly independent.

How do we make an eigenbasis? We choose a basis for each E_{λ_i}, and put these bases together.

Theorem 21.11: Union of Eigenspaces

Let $T : \mathcal{V} \to \mathcal{V}$ be linear, and let $E_{\lambda_1}, \dots, E_{\lambda_k}$ be the distinct eigenspaces.

If we have

- $\beta_1 = \{\vec{\boldsymbol{v}}_{11}, \vec{\boldsymbol{v}}_{12}, \dots, \vec{\boldsymbol{v}}_{1\alpha_1}\}$ = basis for E_{λ_1}
- $\vdots$
- $\beta_k = \{\vec{\boldsymbol{v}}_{k1}, \vec{\boldsymbol{v}}_{k2}, \dots, \vec{\boldsymbol{v}}_{k\alpha_k}\}$ = basis for E_{λ_k}

We then have $\beta = \beta_1 \cup \beta_2 \cup \dots \cup \beta_k$ is a basis in $\mathcal{V}$.

Proof. How do we show that all $\vec{\boldsymbol{v}}_{ij}$ are linearly independent? we have

$$\sum_{i,j} a_{ij} \vec{\boldsymbol{v}}_{ij} = 0$$
$$(a_{11} \vec{\boldsymbol{v}}_{11} + \dots + a_{1\alpha_1} \vec{\boldsymbol{v}}_{1\alpha_1}) + \dots + (a_{k1} \vec{\boldsymbol{v}}_{k1} + \dots + a_{k\alpha_k} \vec{\boldsymbol{v}}_{k\alpha_k} = 0$$

Here, we've just split this up into k parts $\vec{\boldsymbol{w}}_i \in E_{\lambda_i}$ where $\vec{\boldsymbol{w}}_1 + \vec{\boldsymbol{w}}_2 + \dots + \vec{\boldsymbol{w}}_k = 0$. We further can see that $\vec{\boldsymbol{w}}_i$ must all be $\vec{\boldsymbol{0}}$ (they can't be eigenvectors).

Now, for each individual chunk, since each are bases for E_{λ_i}, all coefficients a_{ij} must also be zero. □

Lemma 21.12

We always have $\dim E_{\lambda_1} + \dim E_{\lambda_2} + \dots + \dim E_{\lambda_k} \le n$.

Further, if $\sum_j \dim E_{\lambda_j} = n$, then β is an eigenbasis for $\mathcal{V}$ and T is diagonalizable.

Lemma 21.13

T is diagonalizable if and only if $m_g(\lambda_1) + \dots + m_g(\lambda_k) = n$.

In such a case, if we combine any bases for all E_{λ_j}'s, then we get an eigenbasis for T.

How about algebraic multiplicities?

If T is diagonalizable, then $[T]_\beta$ is the diagonal matrix of n (possibly duplicated) λ_i's where $\lambda_i \in F$.

We know that $\operatorname{char}_T(\lambda)$ splits over F, so we have

$$\begin{aligned}\operatorname{char}_T(\lambda) &= *\lambda_1 - \lambda)(\lambda_2 - \lambda)\cdots(\lambda_n - \lambda) \\ &= (-1)^n (\lambda - \lambda_1)^{m_1} (\lambda - \lambda_2)^{m_2} \cdots)(\lambda - \lambda_k)^{m_k}\end{aligned}$$

In the last expression, λ_i's are all distinct, as we've combined the duplicates.

We then have that $m_1 + m_2 + \dots + m_k = \deg \operatorname{char}_T(\lambda) = n = \dim \mathcal{V}$.

10/18/2021

Lecture 22

Diagonalization III and Applications

What is the relation between $m_a(\lambda_i)$ and $m_g(\lambda_i)$?

Theorem 22.1: Geometric Multiplicity Bounds

If $T : \mathcal{V} \to \mathcal{V}$ is a linear operator, with λ_i as an eigenvalue, then

- $m_g(\lambda_1) \geq 1$ (by definition of eigenvalues)
- $m_g(\lambda_1) \leq m_a(\lambda_1)$

  This last fact is trueeven if $\text{char}_T(\lambda)$ does not split over F, and even if T is not diagonalizable!

Proof. Suppose E_{λ_1} is an eigenspace associated with λ_1. We know that $\dim E_{\lambda_1} = m_g(\lambda_1) = \ell$.

Suppose we create a basis $\beta = \{\vec{\boldsymbol{v}}_1, \dots, \vec{\boldsymbol{v}}_\ell\}$ for E_{λ_1}. We want to extend β to a basis β_1 for $\mathcal{V}$ (by the replacement theorem). That is, $\beta_1 = \{\vec{\boldsymbol{v}}_1, \dots, \vec{\boldsymbol{v}}_\ell, \vec{\boldsymbol{v}}_{\ell+1}, \dots, \vec{\boldsymbol{v}}_n\}$.

Our next question is: what is $[T]_{\beta_1}$. However, we already know that $T(\vec{\boldsymbol{v}}_1) = \lambda_1 \vec{\boldsymbol{v}}_1$, etc. up to $T(\vec{\boldsymbol{v}}_\ell) = \lambda_1 \vec{\boldsymbol{v}}_\ell$. However, we don't know what $T(\vec{\boldsymbol{v}}_{\ell+1})$ through $T(\vec{\boldsymbol{v}}_n)$ map to.

We can say that our matrix of transformation is

$$[T]_{\beta_1} = \left[\begin{array}{c|c} \lambda_1 \mathbf{I}_\ell & * \\ \hline 0 & \mathbf{B} \end{array}\right].$$

We don't know what $\mathbf{B}$ is, nor do we know what the top right block is.

We also want to calculate $[T - \lambda \mathbf{I}_n]_{\beta_1}$; we have

$$\text{char}_T(\lambda) = (\lambda_1 - \lambda)^\ell \cdot g(\lambda).$$

Our question—is it possible that $g(\lambda_1) = 0$? Yes—it is possible that $m_a(\lambda_1) > \ell = m_g(\lambda_1)$.

As such, we've just proven our claim—$m_a(\lambda_1) \geq \ell = m_g(\lambda_1)$. □

Recall:

Lemma 22.2: Criterion 1

$T : \mathcal{V}_F \to \mathcal{V}_F$ is diagonalizable if and only if $m_g(\lambda_1) + \cdots + m - g(\lambda_k) = n$. That is, the sum of all geometric multiplicities for distinct λ_i's add up to n.

Recall:

Lemma 22.3: Criterion 2

If $T : \mathcal{V}_F \to \mathcal{V}_F$ is diagonalizable,

- $\text{char}_T(\lambda)$ splits over F
- $m_g(\lambda_i) = m_a(\lambda_i)$ for all eigenvalues λ_i

(This is in fact an if and only if statement.)

Proof. Using criterion 1, we can prove criterion 2:

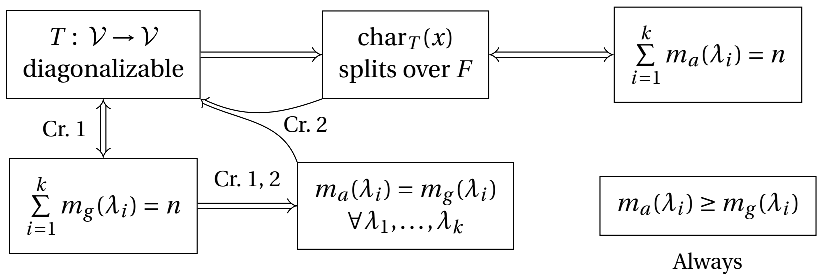
($\implies$) If T is diagonalizable, then we know by criterion 1 that $\sum_{i=1}^k m_g(\lambda_i) = n = \dim \mathcal{V}$. We also know that char_T splits over F; this menas that $\sum_{i=1}^k m_a(\lambda_i) = \dim \mathcal{V}$.

We know that for all λ_i, $m_g(\lambda_i) \leq m_a(\lambda_i)$. If we add all these inequalities up, we have that $\sum_{i=1}^k m_g(\lambda_i) \leq \sum_{i=1}^k m_a(\lambda_i)$. However, we've shown prior that both these sums must be equal to $\dim \mathcal{V}$. As such, all of these inequalities must actually be equalities; $m_g(\lambda_i) = m_a(\lambda_i)$ for all λ_i.

($\Longleftarrow$) Suppose char_T splits over F, and $m_a(\lambda_i) = m_g(\lambda_i)$ for all eigenvalues λ_i.

If the characteristic polynomial splits, then we know that $\sum_{i=1}^k m_a(\lambda_i) = \dim \mathcal{V}$. This is equal to $\sum_{i=1}^k m_g(\lambda_i)$, and by criterion 1, T must be diagonalizable. $\square$

In summary,



22.1 Applications to Recurrences

If we have the Fibonacci numbers $f_{n+1} = f_n + f_{n-1}$ for $n \geq 1$, and $f_0 = 0$, $f_1 = 1$, we want a direct formula for f_{2021}.

We will first paraphrase this problem in terms of linear algebra; let $\vec{\boldsymbol{v}}_n = \begin{bmatrix} f_n \\ f_{n+1} \end{bmatrix} \in \mathbb{R}^2$.

We want to find

$$\vec{\boldsymbol{v}}_{n+1} = \begin{bmatrix} f_{n+1} \\ f_{n+2} \end{bmatrix} = \begin{bmatrix} f_{n+1} \\ f_{n+1} + f_n \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} f_n \\ f_{n+1} \end{bmatrix}.$$

Now we can keep going through the recurrence using this matrix $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$:

$$\vec{\boldsymbol{v}}_n = \mathbf{A}\vec{\boldsymbol{v}}_{n-1} = \mathbf{A}(\mathbf{A}\vec{\boldsymbol{v}}_{n-2}) = \cdots = \mathbf{A}^n \vec{\boldsymbol{v}}_0.$$

So all we need to do is find $\mathbf{A}^n$. If $\mathbf{A}$ is diagonalizable as $\mathbf{A} = \mathbf{QDQ}^{-1}$, then $\mathbf{A}^n = \mathbf{QD}^n\mathbf{Q}^{-1}$, as all of the inverses in the middle will cancel out. The diagonal matrix $\mathbf{D}$ is easy to take powers of, as we just take powers of the entries.

We know that $\text{char}_{\mathbf{A}}(\lambda) = \lambda^2 - \lambda + 1$, so the eigenvalues are $\lambda = \frac{1 \pm \sqrt{5}}{2}$. Since we have two distinct roots, $\mathbf{A}$ is diagonalizable.

We also need the eigenvectors, as they're the columns of $\mathbf{Q}$.

The first eigenvector corresponding to λ_1 is

$$(\mathbf{A} - \lambda_1 \mathbf{I}_2)\vec{\boldsymbol{v}} = \vec{\mathbf{0}} \implies \begin{bmatrix} -\lambda_1 & 1 \\ 1 & 1 - \lambda_1 \end{bmatrix} \vec{\boldsymbol{v}} = \vec{\mathbf{0}} \implies \lambda_1 v_1 = v_2 \implies \vec{\boldsymbol{v}} = \begin{bmatrix} 1 \\ \lambda_1 \end{bmatrix}.$$

Doing the same with λ_2, we have that $\vec{\boldsymbol{v}} = \begin{bmatrix} 1 \\ \lambda_2 \end{bmatrix}$ is the other eigenvector corresponding to λ_2.

We have

$$\mathbf{Q} = \begin{bmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{bmatrix}, \mathbf{Q}^{-1} \frac{1}{\lambda_2 - \lambda_1} \begin{bmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{bmatrix}.$$

However, we know

$$\mathbf{Q} = [\mathcal{I}_{\mathcal{V}}]_\beta^e, \mathbf{Q}^{-1} = [\mathcal{I}_{\mathcal{V}}]_e^\beta.$$

This means our general formula is

$$\begin{aligned}\mathbf{A}^n \begin{bmatrix}0\\1\end{bmatrix} &= \begin{bmatrix}1 & 1\\ \lambda_1 & \lambda_2\end{bmatrix}\begin{bmatrix}\lambda_1^n & 0\\ 0 & \lambda_2^n\end{bmatrix}\frac{1}{\lambda_2-\lambda_1}\begin{bmatrix}\lambda_2 & -1\\ -\lambda_1 & 1\end{bmatrix}\begin{bmatrix}0\\1\end{bmatrix}\\ &= -\frac{1}{\sqrt{5}}\begin{bmatrix}\lambda_1^n & \lambda_2^n\\ \lambda_1^{n+1} & \lambda_2^{n+1}\end{bmatrix}\begin{bmatrix}-1\\1\end{bmatrix}\\ &= \frac{1}{\sqrt{5}}\begin{bmatrix}\lambda_1^n-\lambda_2^n\\ \lambda_1^{n+1}-\lambda_2^{n+1}\end{bmatrix}\end{aligned}$$

This means that

$$f_n = \frac{1}{\sqrt{5}}(\lambda_1^n - \lambda_2^n) = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right).$$

10/20/2021

Lecture 23

Applications of Diagonalization, Systems of ODEs

23.1 Recurrence Relations

Starting to generalize the process of solving a recurrence relation from last time, if we have something like $a_{n+1} = 5a_n - 6a_{n-1}$, we want to convert this into the form

$$\begin{bmatrix}x & y\\ z & t\end{bmatrix}\begin{bmatrix}a_n\\ a_{n+1}\end{bmatrix} = \begin{bmatrix}a_{n+1}\\ a_{n+2}\end{bmatrix}.$$

In the given example, we'd have the coefficient matrix $\mathbf{A} = \begin{bmatrix}0 & 1\\ -6 & 5\end{bmatrix}$. Note that this means that

$$\begin{bmatrix}x & y\\ z & t\end{bmatrix}\begin{bmatrix}a_n\\ a_{n+1}\end{bmatrix} = \begin{bmatrix}xa_n + ya_{n+1}\\ za_n + ta_{n+1}\end{bmatrix} = \begin{bmatrix}0\cdot a_n + 1\cdot a_{n+1}\\ 5\cdot a_{n+1} - 6\cdot a_n\end{bmatrix}.$$

We can see that the coefficient matrix directly corresponds to the constants here; the top row of the coefficient matrix is always a zero and a one.

Suppose we look at $\operatorname{char}_{\mathbf{A}}(\lambda)$.

For the Fibonacci sequence, we have $\operatorname{char}_{\mathbf{A}}(\lambda) = \lambda^2 - \lambda - 1$. In general, with a 2×2 matrix, we have $\operatorname{char}_{\mathbf{A}}(\lambda) = \lambda^2 - \operatorname{tr}\mathbf{A}\cdot\lambda + \det\mathbf{A}$.

For example, in our given example, we have $\operatorname{char}_{\mathbf{A}}(\lambda) = \lambda^2 - (0+5)\lambda + (0\cdot 5 - 1\cdot(-6)) = \lambda^2 - 5\lambda + 6 = \lambda^2 - 5\lambda - (-6)$.

Notice that the coefficients here are exactly the coefficients of the original recurrence relation. This motivates us to ask: is there a way to get $\operatorname{char}_{\mathbf{A}}(\lambda)$ without going through $\mathbf{A}$?

It turns out that there is! A recurrence of the form $a_{n+1} = \alpha a_n + \beta a_{n-1}$ gives the characteristic polynomial $\lambda^2 - \alpha\lambda - \beta = 0$.

Now that we have the characteristic polynomial, we can compute the eigenvalues—a warning here that we need *distinct* eigenvalues; if there are duplicate eigenvalues, the method may be beyond diagonalization.

For the Fibonacci sequence, we have $\lambda_1 = \frac{1+\sqrt{5}}{2} \neq \lambda_2 = \frac{1-\sqrt{5}}{2}$. For our earlier example, we have $\lambda^2 - 5\lambda + 6 = (\lambda-2)(\lambda-3) \implies \lambda_1 = 2 \neq \lambda_2 = 3$.

Lastly, we can observe that we can diagonalize without finding the eigenvectors $\vec{\boldsymbol{v}}_1$ and $\vec{\boldsymbol{v}}_2$!

If $\mathbf{A}$ can be diagonalized, we have $\mathbf{A} = \mathbf{QDQ}^{-1} \implies \mathbf{A}^n = \mathbf{QD}^n\mathbf{Q}^{-1}$. This means that we have

$$\mathbf{A}^n = \begin{bmatrix}* & *\\ * & *\end{bmatrix}\begin{bmatrix}\lambda_1^n & 0\\ 0 & \lambda_2^n\end{bmatrix}\begin{bmatrix}* & *\\ * & *\end{bmatrix},$$

for constant matrices $\mathbf{Q}$ and $\mathbf{Q}^{-1}$. We can simplify this further to see that

$$\mathbf{A}^n = \begin{bmatrix} * & * \\ * & * \end{bmatrix} \begin{bmatrix} *\lambda_1^n & *\lambda_1^n \\ *\lambda_2^n & *\lambda_2^n \end{bmatrix} = \begin{bmatrix} *\lambda_1^n + *\lambda_2^n & *\lambda_1^n + *\lambda_2^n \\ *\lambda_1^n + *\lambda_2^n & *\lambda_1^n + *\lambda_2^n \end{bmatrix}.$$

In the end, we want to compute

$$\mathbf{A}^n \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \mathbf{A}^n \vec{\boldsymbol{v}}_0 = \begin{bmatrix} *\lambda_1^n + *\lambda_2^n \\ *\lambda_1^n + *\lambda_2^n \end{bmatrix}.$$

Notice that a_n is just a linear combination of λ_1^n and λ_2^n; this means that all we need to do is solve for $a_n = a\cdot\lambda_1^n + b\cdot\lambda_2^n$; we'd just substitute the initial conditions and solve the system of equations.

Example 23.1

Given $a_{n+1} = 5a_n - 6a_{n-1}$ and $a_0 = 2020$, $a_1 = 2021$, we have

$$x^2 = 5x - 6 \implies x^2 - 5x + 6 = 0 \implies x_1 = 2,\ x_2 = 3.$$

This means that we have $a_n = a\cdot 2^n + b\cdot 3^n$ for $n \geq 0$. Solving for a and b, we have $a = 4039$, $b = -2019$, giving us a final answer of $a_n = 4039\cdot 2^n - 2019\cdot 3^n$.

This algorithm works for any *linear*, *homogeneous* recurrence relation with *constant* coefficients, where the eigenvalues are distinct roots of the characteristic polynomial.

23.2 Systems of Ordinary Differential Equations

Our goal here is to find functions $x(t) : \mathbb{R} \to \mathbb{R}$ that satisfy a system of differential equations.

In the simple case, suppose we have a single differential equation like the following:

- $x'(t) = x(t) \implies x(t) = c\cdot e^t$
- $x'(t) = 2x(t) \implies x(t) = c\cdot e^{2t}$
- $x'(t) = \lambda x(t) \implies x(t) = c\cdot e^{\lambda t}$

We will try to solve a system of *ordinary, first order* differential equations in this section; that is, equations that are linear, homogeneous, and of order 1 (i.e. only $\frac{\mathrm{d}}{\mathrm{d}x}$).

As an example, suppose we have the system

$$\begin{cases} x_1'(t) = 0\cdot x_1(t) - 2\cdot x_2(t) - 3\cdot x_3(t) \\ x_2'(t) = -1\cdot x_1(t) + 1\cdot x_2(t) - 1\cdot x_3(t) \\ x_3'(t) = 2\cdot x_1(t) + 2\cdot x_2(t) + 5\cdot x_3(t) \end{cases}.$$

Our goal is to turn this system into a system of the form

$$\begin{cases} y_1'(t) = \lambda_1 y_1(t) \\ y_2'(t) = \lambda_2 y_2(t) \\ y_3'(t) = \lambda_3 y_3(t) \end{cases} \implies \begin{cases} y_1(t) = c_1 e^{\lambda_1 t} \\ y_2(t) = c_2 e^{\lambda_2 t} \\ y_3(t) = c_3 e^{\lambda_3 t} \end{cases}.$$

This kind of system is called an "uncoupled" system, and each function can be solved for directly. Here, we can see that $x_i(t)$'s must all be some linear combination of the $y_i(t)$'s, and we can solve backwards from here.

Firstly, we can establish some notation; in the prior example, we have the coefficient matrix $\mathbf{A}$, with $\vec{\boldsymbol{x}}(t)$ and $\vec{\boldsymbol{x}}'(t)$ as vector functions:

$$\mathbf{A} = \begin{bmatrix} 0 & -2 & -3 \\ -1 & 1 & -1 \\ 2 & 2 & 5 \end{bmatrix},\quad \vec{\boldsymbol{x}}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix},\quad \vec{\boldsymbol{x}}'(t) = \begin{bmatrix} x_1'(t) \\ x_2'(t) \\ x_3'(t) \end{bmatrix}.$$

Paraphrased, our system is now $\mathbf{A}\vec{\boldsymbol{x}}(t) = \vec{\boldsymbol{x}}'(t)$.

Now we will turn to a procedure to solve such systems of differential equations. If **A** is diagonalizable, we have that $\mathbf{A} = \mathbf{QDQ}^{-1}$, where **D** is diagonal. This means that we have

$$\mathbf{A}\vec{\boldsymbol{x}}(t) = \mathbf{QDQ}^{-1}\vec{\boldsymbol{x}}(t) = \vec{\boldsymbol{x}}'(t) \implies \mathbf{D}\underbrace{\mathbf{Q}^{-1}\vec{\boldsymbol{x}}(t)}_{\vec{\boldsymbol{y}}(t)} = \underbrace{\mathbf{Q}^{-1}\vec{\boldsymbol{x}}'(t)}_{\vec{\boldsymbol{y}}'(t)} \implies \mathbf{D}\vec{\boldsymbol{y}}(t) = \vec{\boldsymbol{y}}'(t).$$

At this point, we've converted our system of differential equations into our desired form; the diagonal matrix completely uncouples the equations into

$$\mathbf{D}\vec{\boldsymbol{y}}(t) = \vec{\boldsymbol{y}}'(t) \implies \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \end{bmatrix} = \begin{bmatrix} y_1'(t) \\ y_2'(t) \\ y_3'(t) \end{bmatrix} \implies \begin{cases} y_1'(t) = \lambda_1 y_1(t) \\ y_2'(t) = \lambda_2 y_2(t) \\ y_3'(t) = \lambda_3 y_3(t) \end{cases}.$$

Since we can solve for $\vec{\boldsymbol{y}}(t)$ easily here, we can also solve for $\vec{\boldsymbol{x}}(t) = \mathbf{Q}\vec{\boldsymbol{y}}(t)$.

However, we need both eigenvalues and eigenvectors; recall that **Q** is our matrix of eigenvectors, i.e. $\mathbf{Q} = [\mathcal{I}_\mathcal{V}]_\beta^e$. Further, notice that $\mathbf{Q}^{-1} = [\mathcal{I}_\mathcal{V}]_e^\beta$ is not needed; we do not need to explicitly change basis, and all we need to do is change back to solve for $\vec{\boldsymbol{x}}(t)$ using **Q** only.

Note that if **A** is not diagonalizable, the method of solving $\mathbf{A}\vec{\boldsymbol{x}}9t) = \vec{\boldsymbol{x}}'(t)$ is beyond diagonalization and requires the Jordan Canonical Form.

10/22/2021

Lecture 24

Invariant and Cyclic subspaces, Cayley-Hamilton Theorem

24.1 Invariant Subspaces

Definition 24.1: T-invariance

Suppose we have a linear operator $T : \mathcal{V} \to \mathcal{V}$, and let $\mathcal{W} \leqq \mathcal{V}$ (i.e. $\mathcal{W}$ is a subspace of $\mathcal{V}$). Then, $\mathcal{W}$ is T-invariant if $\vec{\boldsymbol{w}} \in \mathcal{W} \implies T(\vec{\boldsymbol{w}}) \in \mathcal{W}$.

That is, *invariant* means that $\mathcal{W}$ doesn't move (change) under T: $T(\mathcal{W}) \subseteq \mathcal{W}$.

Example 24.2

Suppose we take a reflection $T : \mathbb{R}^2 \to \mathbb{R}^2$ across the x-axis: $\begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} x \\ -y \end{bmatrix}$.

The subspace $\mathcal{W}_1$ consisting of the x-axis is $\{c \cdot \vec{\boldsymbol{e}}_1 \mid c \in \mathbb{R}\} = E_1$; it's the eigenspace corresponding to the eigenvector 1.

The subspace $\mathcal{W}_2$ consisting of the y-axis is $\{c \cdot \vec{\boldsymbol{e}}_2 \mid c \in \mathbb{R}\} = E_{-1}$; it's the eigenspace corresponding to the eigenvector -1.

Both of these subspaces are T-invariant; the x-axis gets mapped to itself, whereas the y-axis gets mapped onto itself through its reflection—vectors still stay on the y-axis after applying T.

The entire plane $\mathbb{R}^2 = \mathcal{V}$ itself is also T-invariant, as well as the zero vector space $\{\vec{\mathbf{0}}\}$.

Lemma 24.3

Suppose we have a linear operator $T: \mathcal{V} \to \mathcal{V}$. The following subspaces are always T-invariant: $\{\vec{\mathbf{0}}\}$, $\mathcal{V}$, E_λ, $\operatorname{Ker} T = E_0$, $\operatorname{Im} T$

Proof. The zero vector space $\{\vec{\mathbf{0}}\}$ and $\mathcal{V}$ are both trivially T-invariant; the former will map to itself, and the latter is where all vectors must come from and map to.

If $\mathcal{W} = E_\lambda$, we can show that $\vec{\boldsymbol{w}} \in E_\lambda$ implies that $T(\vec{\boldsymbol{w}}) \in E_\lambda$ as well; we have $T(\vec{\boldsymbol{w}}) = \lambda \vec{\boldsymbol{w}}$, which is a scalar multiple of $\vec{\boldsymbol{w}}$ itself, and as such must be in E_λ. This proves the claim for $\operatorname{Ker} T = E_0$ as well.

If $\mathcal{W} = \operatorname{Im} T$, we can show that $\vec{\boldsymbol{w}} \in \operatorname{Im} T$ implies that $T(\vec{\boldsymbol{w}}) \in \operatorname{Im} T$ as well; by definition, any output of T must be in its image. As such, $T(\vec{\boldsymbol{w}}) \in \operatorname{Im} T$. □

Example 24.4

Suppose we have the linear operator T with the following matrix of transformation:

$$[T]_\beta = \begin{bmatrix} 1 & 2 & -1 \\ 3 & 4 & 6 \\ 0 & 0 & 5 \end{bmatrix}.$$

Notice that if $\beta = \{\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2, \vec{\boldsymbol{v}}_3\}$, we have

$$\begin{aligned} T(\vec{\boldsymbol{v}}_1) &= \vec{\boldsymbol{v}}_1 + 3\vec{\boldsymbol{v}}_3 \\ T(\vec{\boldsymbol{v}}_2) &= 2\vec{\boldsymbol{v}}_1 + 5\vec{\boldsymbol{v}}_2 \\ T(\vec{\boldsymbol{v}}_3) &= -\vec{\boldsymbol{v}}_1 + 6\vec{\boldsymbol{v}}_2 + 5\vec{\boldsymbol{v}}_3 \end{aligned}$$

Here, $T(\vec{\boldsymbol{v}}_1)$ and $T(\vec{\boldsymbol{v}}_2)$ are both linear combinations of $\vec{\boldsymbol{v}}_1$ and $\vec{\boldsymbol{v}}_2$, and thus $\mathcal{W} = \operatorname{span}\left\{\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2\right\}$ is T-invariant.

Further, notice that this happens because the first two columns of $[T]_\beta$ have zeroes in the third row; the images are restricted to some linear combination of the input basis vectors.

In general, how do we locate some T-invariant subspaces? Suppose we have something of the form

$$[T]_\beta = \begin{bmatrix} * & 0 & * \\ * & \mathbf{A} & * \\ * & 0 & * \end{bmatrix}.$$

Specifically, suppose we have a block matrix $\mathbf{A}$ where everything above and below are zeroes, where the columns that $\mathbf{A}$ resides in correspond to vectors $\{\vec{\boldsymbol{v}}_i, \vec{\boldsymbol{v}}_{i+1}, \dots, \vec{\boldsymbol{v}}_j\}$. Then, $\operatorname{span}\left\{\vec{\boldsymbol{v}}_i, \vec{\boldsymbol{v}}_{i+1}, \dots, \vec{\boldsymbol{v}}_j\right\}$ is a T-invariant subspace.

Further, suppose we think of T restricted on the T-invariant subspace $\mathcal{W}$; that is, $T_{\mathcal{W}}: \mathcal{W} \to \mathcal{W}$. We then have $[T_{\mathcal{W}}]_{\beta_1} = \mathbf{A}$.

24.2 Cyclic Subspaces

Definition 24.5: T-Cyclic Subspace

Suppose we have a linear operator $T: \mathcal{V} \to \mathcal{V}$, with $\vec{\boldsymbol{v}} \in \mathcal{V}$. Then the T-cyclic subspace generated by $\vec{\boldsymbol{v}}$ is

$$\operatorname{span}\left\{\vec{\boldsymbol{v}}, T(\vec{\boldsymbol{v}}), T^2(\vec{\boldsymbol{v}}), \dots, T^k(\vec{\boldsymbol{v}}), \dots\right\}.$$

10/25/2021

Lecture 25

Jordan Canonical Form

Let us extend the lemma from last time; let $\mathcal{W} = \operatorname{span}\{\vec{v}, T(\vec{v}), T^2(\vec{v}), \dots, T^{k-1}(\vec{v})\}$ is T-cyclic with basis $\beta = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$. Let $T^k(\vec{v}) = a_0\vec{v} + a_1 T(\vec{v}) + a_2 T^2(\vec{v}) + \cdots + a_{k-1}T^{k-1}(\vec{v})$.

Here, we have

$$[T]_\beta = \begin{bmatrix} 0 & 0 & \cdots & 0 & a_0 \\ 1 & 0 & \cdots & 0 & a_1 \\ 0 & 1 & \cdots & 0 & a_2 \\ 0 & 0 & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & a_{k-1} \end{bmatrix}.$$

Recall that we can directly find the characteristic polynomial:

$$\operatorname{char}_T(\lambda) = (-1)^k\left(a_0 + a_1\lambda + a_2\lambda^2 + \cdots + a_{k-1}\lambda^{k-1} + \lambda^k\right).$$

T-cyclic spaces are easy to work with; we can easily find a basis, the transformation matrix, the characteristic polynomial, the dimension of the subspace, etc.

25.1 Jordan Canonical Form

A variation of the idea of T-cyclic spaces are the so-called "Jordan blocks", which are a special type of T-cyclic spaces.

$$\mathbf{J}_n(\alpha) = \mathbf{A} = \begin{bmatrix} \alpha & 1 & 0 & \cdots & 0 \\ 0 & \alpha & 1 & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \alpha & 1 \\ 0 & \cdots & 0 & 0 & \alpha \end{bmatrix}.$$

Definition 25.1: Jordan Block

A Jordan block corresponding to α of size $n \times n$ satisfies

- Upper triangular matrix
- Diagonal entries are $a_{ii} = \alpha$
- Above diagonal $a_{i,i+1} = 1$
- All other $a_{ij} = 0$.

We will view this in terms of linear transformations. We have a linear operator $T : \mathcal{V} \to \mathcal{V}$. In some basis $\beta = \{\vec{v}_1, \dots, \vec{v}_n\}$, we have $[T]_\beta = \mathbf{A}$ from above. What does T do to these vectors?

$$\begin{aligned} T(\vec{v}_1) &= \alpha\vec{v}_1 \\ T(\vec{v}_2) &= \vec{v}_1 + \alpha\vec{v}_2 \\ T(\vec{v}_3) &= \vec{v}_2 + \alpha\vec{v}_3 \\ &\vdots \\ T(\vec{v}_n) &= \vec{v}_{n-1} + \alpha\vec{v}_n \end{aligned}$$

Further, we can directly write the characteristic polynomial—the matrix is upper triangular:

$$\mathrm{char}_{\mathbf{A}}(\lambda) = (\alpha - \lambda)^n.$$

Here, notice that $m_a(\alpha) = n$; it is maximal. Further, $m_g(\alpha) = 1$; it is minimal. We can see this with the following:

$$(\mathbf{A} - \alpha\mathbf{I})\vec{\boldsymbol{v}} = \vec{\mathbf{0}} \implies \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \vec{\boldsymbol{v}} = \vec{\mathbf{0}} \implies v_1 = t,\ v_i = 0,\ i > 1.$$

Notice that this means that the dimension of the eigenspace is 1. As such, $\mathbf{A} = \mathbf{J}_n(\alpha)$ is not diagonalizable unless $n = 1$.

What is our goal here? We want the Jordan matrix, consisting of Jordan blocks along the diagonal, and zeroes elsewhere:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & 0 & \cdots & 0 & 0 \\ 0 & \mathbf{A}_2 & 0 & \ddots & 0 \\ \vdots & 0 & \ddots & 0 & \vdots \\ 0 & \ddots & 0 & \mathbf{A}_{k-1} & 0 \\ 0 & 0 & \cdots & 0 & \mathbf{A}_k \end{bmatrix}.$$

Each $\mathbf{A}_i$ here is a Jordan block; this means that $\mathbf{A}$ is upper triangular with non-zero entries only (possbly) along the diagonal and one diagonal above it.

We will call $[T]_\beta = \mathbf{A}$ the Jordan canonical form of T, with β being the Jordan basis.

Because this matrix is upper triangular, we can just take the entries along the diagonal.

$$\mathrm{char}_T(\lambda) = (-1)^n (\lambda - a_1)^{m_1} \cdots (\lambda - a_k)^{m_k}.$$

Further, this means that $\mathrm{char}_T(\lambda)$ splits over F (necessarily!)

Theorem 25.2: Jordan's Theorem

Let $T : \mathcal{V} \to \mathcal{V}$ be a linear operator such that $\mathrm{char}_T(\lambda)$ splits over F. Then there exists a Jordan canonical basis β in which T has a Jordan canonical form (JCF).

Further, the JCF is unique up to reordering of the Jordan blocks, but the basis is not unique (trivially, we can rescale the first eigenvector, but there are others).

Example 25.3

Suppose $\mathbf{A} = \mathbf{I}_n$. Any basis β is a Jordan basis for $\mathbf{I}_n$ (as any basis will give us the identity). However, the Jordan form will always be $\mathbf{I}_n$.

Suppose we have $\mathbf{A} = \mathbf{J}_n = [T]_\beta$. Suppose we take $\mathbf{A} - \alpha\mathbf{I}$ (there is only one eigenvalue associated with a Jordan block). This gives us

$$\mathbf{B} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

A followup question: what is $\mathbf{B}^2$? We get zeroes along the diagonal and the diagonal above, but ones along the diagonal above that:

$$\mathbf{B}^2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & \ddots & \ddots & 1 \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 0 \end{bmatrix}.$$

For example, suppose we have $\mathbf{A}$ be a 4×4 matrix $\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. We have

$$\mathbf{A}^2 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
$$\mathbf{A}^3 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$
$$\mathbf{A}^4 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Why is this happening? Let us look at the Cayley-Hamilton Theorem; we have $\mathrm{char}_{\mathbf{A}}(\lambda) = \lambda^4$. This means that $\mathrm{char}_{\mathbf{A}}(\mathbf{A}) = \mathbf{A}^4 = \mathbf{0}$.

But we haven't proven Cayley-Hamilton yet; let us offer another explanation involving linear transformations.

We have $U = T - \alpha\mathcal{I}_{\mathcal{V}} : \mathcal{V} \to \mathcal{V}$. We saw what T did to the basis $\beta = \{\vec{\boldsymbol{v}}_1, \dots, \vec{\boldsymbol{v}}_n\}$; but what does U do to this basis?

$$\begin{aligned} U(\vec{\boldsymbol{v}}_1) &= \vec{\mathbf{0}} \\ U(\vec{\boldsymbol{v}}_2) &= \vec{\boldsymbol{v}}_2 \\ U(\vec{\boldsymbol{v}}_3) &= \vec{\boldsymbol{v}}_3 \\ &\vdots \\ U(\vec{\boldsymbol{v}}_n) &= \vec{\boldsymbol{v}}_{n-1} \end{aligned}$$

As such, each application of U moves the basis vectors to the right, eventually giving us all zeroes.

Our conclusion here is that $U^n = (T - \alpha\mathcal{I})^n$ sends everything to zero. That is, $\mathcal{V} = \mathrm{Ker}(T - \alpha\mathcal{I})^n$ where $n = \dim\mathcal{V}$.

If we generalize this process, suppose $T : \mathcal{V} \to \mathcal{V}$ be a linear operator and $[T]_\beta$ is the Jordan Canonical Form in the Jordan basis β, where all the Jordan blocks $\mathbf{A}_1, \dots, j\mathbf{A}_k$ correspond to the same eigenvalue α. (This is an intermediate situation; not quite completely general.)

We have $[T - \alpha\mathcal{I}]_\beta$ changes the diagonal into zeroes, leaving the ones along the diagonal above (but still zero between Jordan blocks).

What happens if we raise this matrix to the power of n? We can see that each one of the 1's move up with each successive power, and as such $(T - \alpha\mathcal{I})^n = 0_{\mathcal{V}}$, i.e. $(T - \alpha\mathcal{I})^n \vec{\boldsymbol{v}} = \vec{\mathbf{0}}$.

What are $m_a(\alpha$ and $m_g(\alpha)$? We know that $m_a(\alpha) = n$, while $m_g(\alpha) = k$, i.e. the geometric multiplicity is the number of Jordan blocks corresponding to α; we have one eigenvector per Jordan block.

Hence, $\dim E_\alpha = k$; T is not diagonalizable if $k < n$, but it will be diagonalizable if $k = n$. That is, T is not diagonalizable as long as some Jordan block is of size $> 1 \times 1$.

Equivalently,

Lemma 25.4

T is diagonalizable if and only if all Jordan blocks are 1×1; the JCF is diagonal itself!

10/27/2021

Lecture 26

Generalized Eigenspaces

Definition 26.1: Generalized Eigenspace

Suppose we have a linear operator $T : \mathcal{V}_F \to \mathcal{V}_F$, and $\alpha \in F$. The generalized eigenspace corresponding to α is the set $K_\alpha(T) = \{\vec{\boldsymbol{v}} \in \mathcal{V} \mid (T - \alpha \mathbf{I})^m \vec{\boldsymbol{v}} = \vec{\mathbf{0}},\ m \in \mathbb{N}\}$.

In other words, we look at the set of all vectors that are sent to zero by some power of $T - \alpha\mathbf{I}$; any $\vec{\boldsymbol{v}} \in K_\alpha(T)$ is called a *generalized eigenvector* of T corresponding to α.

Example 26.2

We can see that $E_\alpha = \operatorname{Ker}(T - \alpha\mathcal{I})^1$.

This is because $\vec{\boldsymbol{v}} \in E_\alpha$ iff $T(\vec{\boldsymbol{v}}) = \alpha\vec{\boldsymbol{v}}$. This means that $T(\vec{\boldsymbol{v}}) - \alpha\vec{\boldsymbol{v}} = \vec{\mathbf{0}}$, and $(T - \alpha\mathbf{I})\vec{\boldsymbol{v}} = \vec{\mathbf{0}}$, which means that $\vec{\boldsymbol{v}} \in \operatorname{Ker}(T - \alpha\mathcal{I})^1$.

Our conclusion is that $E_\alpha \subseteq K_\alpha(T)$. In this case, we have $m = 1$.

26.1 Properties of Generalized Eigenspaces

Is $K_\alpha(T)$ a subspace of $\mathcal{V}$? Is it T-invariant? Is it T-cyclic? It turns out that $K_\alpha(T)$ is a subspace and also T-invariant.

Further, can we break up $\mathcal{V}$ into building blocks $K_\alpha(T)$ for distinct eigenvalues $\alpha_1, \dots, \alpha_k$ of T? That is, are we able to have

$$\mathcal{V} = K_{\alpha_1}(T) \oplus K_{\alpha_2}(T) \oplus \cdots \oplus K_{\alpha_k}(T)?$$

Lemma 26.3

$K_\alpha(T) \leqq \mathcal{V}$.

Proof. We know that $\vec{\mathbf{0}} \in K_\alpha(T)$, because $\vec{\mathbf{0}}$ will be set to $\vec{\mathbf{0}}$ for anything, say $m = 1$.

Suppose we haev $\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2 \in K_\alpha(T)$ and $a_1, a_2 \in F$. We want to know $a_1\vec{\boldsymbol{v}}_1 + a_2\vec{\boldsymbol{v}}_2$.

We know that $(T - \alpha\mathbf{I})^{m_1}\vec{\boldsymbol{v}}_1 = \vec{\mathbf{0}}$, and $(T - \alpha\mathbf{I})^{m_2}\vec{\boldsymbol{v}}2 = \vec{\mathbf{0}}$. We can see that the largest of the two powers must kill both; that is, let $m = \max(m_1, m_2)$.

We have that

$$(T - \alpha\mathbf{I})^m(a_1\vec{\boldsymbol{v}}1 + a_2\vec{\boldsymbol{v}}_2) = a_1(T - \alpha\mathbf{I})^m\vec{\boldsymbol{v}}_1 + a_2(T - \alpha\mathbf{I})^m\vec{\boldsymbol{v}}_2 = \vec{\mathbf{0}}.$$

Since $m \geq m_1$ and $m \geq m_2$, both powers will kill $\vec{\boldsymbol{v}}_1$ and $\vec{\boldsymbol{v}}2$, meaning we're summing two zeroes and we get the zero vector.

As such, since $K_\alpha(T)$ is closed under linear combinations as well, it is indeed a vector subspace of $\mathcal{V}$. $\square$

Lemma 26.4

$K_\alpha(T)$ is T-invariant.

Proof. We take $\vec{\boldsymbol{v}} \in K_\alpha(T)$, and we want to show that $T(\vec{\boldsymbol{v}}) \in K_\alpha(T)$.

We know that $(T - \alpha\mathbf{I})^m \vec{\boldsymbol{v}} = \vec{\mathbf{0}}$. We can observe that the same power m will also kill $T(\vec{\boldsymbol{v}})$.

We can first see that $(T-\alpha\mathbf{I})^m$ and T commute; we can observe the case for $m = 1$ and prove the rest by induction (omitted here). This means that $(T - \alpha\mathbf{I})^m T(\vec{\boldsymbol{v}}) = T(T - \alpha\mathbf{I})^m \vec{\boldsymbol{v}} = \vec{\mathbf{0}}$.

As such, $T(\vec{\boldsymbol{v}}) \in K_\alpha(T)$, and $K_\alpha(T)$ is T-invariant. $\square$

Lemma 26.5

$K_\alpha(T)$ is not necessarily T-cyclic.

Proof. Suppose $T = \mathcal{I}_\mathcal{V}$. T has only one eigenvalue 1. We have that $K_1(\mathcal{I}_\mathcal{V}) = E_1 = \mathcal{V}$, which is not T-cyclic if $\dim \mathcal{V} \geq 2$. $\square$

Suppose $\mathcal{W}_1 = K_\alpha(T)$. Our claim here is that $T_{\mathcal{W}_1} : \mathcal{W}_1 \to \mathcal{W}_1$ has only one eigenvalue $\lambda = \alpha$.

Proof. Suppose $\mu \neq \alpha$ is an eigenvalue of $T_{K_\alpha(T)} = \mathcal{W}_1$. This means that there exists a nonzero vector $\vec{\boldsymbol{v}} \in \mathcal{W}_1$ such that $T(\vec{\boldsymbol{v}}) = \mu\vec{\boldsymbol{v}}$. Further, by definition, we have $(T - \alpha\mathbf{I})^m \vec{\boldsymbol{v}} = \vec{\mathbf{0}}$, for some $m \geq 1$.

This means that $\vec{\boldsymbol{v}}$ is killed by two different things. The problem arises if we have $(T - \alpha\mathbf{I})\vec{\boldsymbol{v}}$; we get

$$(T - \alpha\mathbf{I})\vec{\boldsymbol{v}} = \mu\vec{\boldsymbol{v}} - \alpha\vec{\boldsymbol{v}} = (\mu - \alpha)\vec{\boldsymbol{v}}.$$

This means that $(T - \alpha\mathbf{I})^m \vec{\boldsymbol{v}} = (\mu - \alpha)^m \vec{\boldsymbol{v}} = \vec{\mathbf{0}}$. We know that $\vec{\boldsymbol{v}} \neq \vec{\mathbf{0}}$, so as such $\mu - \alpha = 0$, and we have arrived at a contradiction; we must have $\mu = \alpha$. $\square$

Lemma 26.6

If $\alpha \neq \beta$ then $K_\alpha(T) \cap K_\beta(T) = \{\vec{\mathbf{0}}\}$.

Recall that we've already shown that $E_\alpha \cap E_\beta = \{\vec{\mathbf{0}}\}$.

Proof. Suppose we have $\vec{\boldsymbol{v}}$ is a nonzero vector such that $\vec{\boldsymbol{v}} \in K_\alpha(T)$ and $\vec{\boldsymbol{v}} \in K_\beta(T)$.

This means that $(T - \alpha\mathbf{I})^{m_1} \vec{\boldsymbol{v}} = \vec{\mathbf{0}}$, and $(T - \beta\mathbf{I})^{m_2} \vec{\boldsymbol{v}} = \vec{\mathbf{0}}$, where $m_1, m_2 \geq 1$ are the first powers that kill $\vec{\boldsymbol{v}}$. WLOG, suppose $\min(m_1, m_2) = m_1$.

Let $H_\alpha = (T - \alpha\mathbf{I})$, and let $H_\alpha[\vec{\boldsymbol{v}}] = \operatorname{span}\left\{\vec{\boldsymbol{v}}, H_\alpha\vec{\boldsymbol{v}}, H_\alpha^2\vec{\boldsymbol{v}}, \dots, H_\alpha^{m_1-1}\vec{\boldsymbol{v}}\right\}$. We know that at some point we will arrive at $\vec{\mathbf{0}}$, so we will end at $H_\alpha^{m_1-1}\vec{\boldsymbol{v}}$.

Suppose we let $\vec{\boldsymbol{v}}_1 = H_\alpha^{m_1-1}\vec{\boldsymbol{v}}$. We know that $H_\alpha\vec{\boldsymbol{v}}_1 = H_\alpha^{m_1}\vec{\boldsymbol{v}} = \vec{\mathbf{0}}$. As such, $(T - \alpha\mathbf{I})\vec{\boldsymbol{v}}_1 = \vec{\mathbf{0}}$, and $T\vec{\boldsymbol{v}}_1 = \alpha\vec{\boldsymbol{v}}_1$. As such, $\vec{\boldsymbol{v}}_1 \in E_{\alpha,T}$; $\vec{\boldsymbol{v}}_1$ is an eigenvector of T corresponding to eigenvalue α.

Now, let us take $H_\alpha^{m_1-1} H_\beta^{m_2}\vec{\boldsymbol{v}}$. What will happen? We can see that $H_\alpha^{m_1-1}$ and $H_\beta^{m_2}$ commute (we can again use induction here by starting with just the first powers). As such,

$$H_\alpha^{m_1-1} H_\beta^{m_2}\vec{\boldsymbol{v}} = H_\beta^{m_2} H_\alpha^{m_1-1}\vec{\boldsymbol{v}} = \vec{\mathbf{0}}.$$

However, note that $H_\alpha^{m_1-1}\vec{\boldsymbol{v}} = \vec{\boldsymbol{v}}_1$; this means that $H_\beta^{m_2}\vec{\boldsymbol{v}}_1 = \vec{\mathbf{0}}$. This means that $(T-\beta\mathbf{I})\vec{\boldsymbol{v}}_1 = (\alpha-\beta)\vec{\boldsymbol{v}}_1$, and $H_\beta^{m_2} = (T-\beta\mathbf{I})^{m_2}\vec{\boldsymbol{v}}_1 = (\alpha-\beta)^{m_2}\vec{\boldsymbol{v}}_1$.

This is our contradiction; we know that $\alpha-\beta\neq 0$, and we also know that $\vec{\boldsymbol{v}}_1 \neq 0$. This means that our very original assumption that $\vec{\boldsymbol{v}} \in K_\alpha(T)\cap K_\beta(T)$ must be false; the intersection must be trivial. □

How does this all fit in with the Jordan Canonical Form? If we have distinct eigenvalues $\alpha_1,\dots,\alpha_k$ of T, we can see that $K_{\alpha_1}(T)$, $K_{\alpha_2}(T)$, $\dots K_{\alpha_k}(T)$ are all disjoint, and their direct sum is $\mathcal{V}$. (This will be proven next time.)

Since each $K_{\alpha_i}(T)$ is T-invariant, the union of the bases for each $K_{\alpha_i}(T)$ must be a basis for $\mathcal{V}$.

Further, in this basis β,

$$[T]_\beta = \begin{bmatrix} [T_{K_{\alpha_1}(T)}]_{\beta_1} & 0 & \cdots & 0 \\ 0 & [T_{K_{\alpha_2}(T)}]_{\beta_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & [T_{K_{\alpha_k}(T)}]_{\beta_k} \end{bmatrix}.$$

This is one of our main steps in finding the Jordan Canonical Form.

10/29/2021

Lecture 27

Direct Sums of Generalized Eigenspaces

Does m in the definition of $K_\lambda(T)$ really depend on each $\vec{\boldsymbol{v}}$ or is it independent of the $\vec{\boldsymbol{v}}$'s?

Recall that if we have $U = T - \lambda\mathcal{I}\colon \mathcal{V}\to\mathcal{V}$, each application of U will decrease the size of the image. (This happens for any transformation!)

This means that we have

$$\mathcal{V} \supset \operatorname{Im} U \supset \operatorname{Im} U^2 \supset \cdots \supset U^N = \operatorname{Im} U^{N+1} = \operatorname{Im} U^{N+2} = \cdots.$$

At some point, we are guaranteed that this process stabilizes; and after the first time the image does not change size, U becomes an isomorphism on this smaller image, and thus stabilizes on every other successive application of U.

Equivalently, the kernel will increase in size, and will eventually stabilize *at the same* N.

$$\{\vec{\mathbf{0}}\} \subset \operatorname{Ker} U \subset \operatorname{Ker} U^2 \subset \cdots \subset \operatorname{Ker} U^N = \operatorname{Ker} U^{N+1} = \operatorname{Ker} U^{N+2}.$$

By the dimension theorem, we can see that $\dim\operatorname{Ker} U^k = \dim ImageU^k = \dim\mathcal{V}$ for each k. That is, $\dim\operatorname{Ker} U^k$ and $\dim\operatorname{Im} U^k$ stabilize simultaneously!

With this in mind, we can redefine $K_\lambda(T)$; we originally have $\{\vec{\boldsymbol{v}}\in\mathcal{V} \mid (T-\lambda\mathcal{I})\vec{\boldsymbol{v}} = \vec{\mathbf{0}} \text{ for some } m\in\mathbb{N}\}$.

With our previous stabilization ideas in mind, we have

$$K_\lambda(T) = \operatorname{Ker}(T-\lambda\mathcal{I})^1 \cap \operatorname{Ker}(T-\lambda\mathcal{I})^2 \cap \cdots = \operatorname{Ker} U \subset \operatorname{Ker} U^2 \subset \operatorname{Ker} U^3 \subset \cdots \subset \operatorname{Ker} U^N = \operatorname{Ker} U^{N+1} = \boxed{\operatorname{Ker} U^N}.$$

Because $\operatorname{Ker} U^N$ is the largest kernel that we could possibly get, this must contain all generalized eigenvectors.

As such, we have

Definition 27.1: Generalized Eigenspace (alternative definition)

A generalized eigenspace $K_\lambda(T)$ is the kernel of some super-Halloween operator $U^N = S$.

$$K_\lambda(T) = \operatorname{Ker} U^N = \operatorname{Ker}(T - \lambda\mathcal{I})^N,$$

for some large enough (fixed) integer N depending on T and λ.

Here, N is the *stabilizing constant*.

27.1 Splitting $\mathcal{V}$ as a Direct Sum

As a setup, we have $S = \operatorname{Ker}(T - \lambda\mathcal{I})^N : \mathcal{V} \to \mathcal{V}$. This operator is the first efficient super-Halloween operator associated to λ.

We would like to build a basis; let $\beta_1 = \{\vec{\boldsymbol{v}}1, \vec{\boldsymbol{v}}_2, \ldots \vec{\boldsymbol{v}}_k\}$ be a basis for $\operatorname{Ker} S$, which is equal to $K_\lambda(T)$ by definition. Further, let us create $\beta = \{\vec{\boldsymbol{v}}_{k+1}, \ldots, \vec{\boldsymbol{v}}_n\}$ be a basis for $\Im S$.

Theorem 27.2: Direct Sum of Kernel and Image of S

Our claim is that $\beta = \beta_1 \cup \beta_2$ is a basis for $\mathcal{V}$. This then immediately tells us that $\mathcal{V} = \operatorname{Ker} S \oplus \operatorname{Im} S$.

Proof. We know that $\dim \operatorname{Ker} S + \dim \operatorname{Im} S = \dim \mathcal{V} = n$. Since we already know that we have the correct number of vectors, it suffices to show that they are linearly independent or that they span $\mathcal{V}$ (it is not necessary to show both).

It turns out that it is easier to show that they are linearly independent; we already know that these vectors are linearly independent within their own bases, so we need toss how that the two together are still linearly independent.

If $\operatorname{span}\{\beta_1\}$ and $\operatorname{span}\{\beta_2\}$ intersect with $\{\vec{\mathbf{0}}\}$, then we're done; it is impossible for there to be some linear dependence between the spaces (as the linear dependence would mean that some vector is in both spans).

Suppose we look at

$$\begin{aligned} a_1\vec{\boldsymbol{v}}_1 + \ldots + a_k\vec{\boldsymbol{v}}_k + a_{k+1}\vec{\boldsymbol{v}}_{k+1} + \cdots + a_n\vec{\boldsymbol{v}}_n &= \vec{\mathbf{0}} \\ a_1S(\vec{\boldsymbol{v}}_1) + \cdots + a_kS(\vec{\boldsymbol{v}}_k) + a_{k+1}S(\vec{\boldsymbol{v}}_{k+1}) + \cdots + a_nS(\vec{\boldsymbol{v}}_n) &= \vec{\mathbf{0}} \\ a_{k+1}S(\vec{\boldsymbol{v}}_{k+1}) + \cdots + a_nS(\vec{\boldsymbol{v}}_n) &= \vec{\mathbf{0}} && (\text{first } k \text{ vectors} \in \operatorname{Ker} S) \end{aligned}$$

If all of $a_{k+1}, \ldots a_n$ are linearly independent, then we would have some linear relation between the first k basis vectors (i.e. the basis for $\operatorname{Ker} S$), and thus all coefficients must be zero, and we're done.

As such, we must have that $\{S(\vec{\boldsymbol{v}}_{k+1}, \ldots, S(\vec{\boldsymbol{v}}_n)\}$ are linearly dependent. However, we know that $\operatorname{Im} S = \operatorname{span}\{\vec{\boldsymbol{v}}_{k+1}, \ldots, \vec{\boldsymbol{v}}_n\}$.

We further know that $\operatorname{Im} S^2 = \operatorname{span}\{S(\vec{\boldsymbol{v}}_{K+1}, \ldots, S(\vec{\boldsymbol{v}}_n))\}$. But since we've established that these RHS vectors are linearly dependent, this span must be at least one dimension less than $\operatorname{Im} S$.

This is impossible—we've defined S to be at the point where the image does not decrease in size after additional applications of $U = (T - \lambda\mathcal{I})$, but we've found that $\Im S^2 = \Im U^{2N}$ decreased in size.

As such, $\{S(\vec{\boldsymbol{v}}_{k+1}, \ldots, S(\vec{\boldsymbol{v}}_n)\}$ must be linearly independent, and this forces all the coefficients $a_1, \ldots a_n$ to be zero.

Thus, the vectors $\{\vec{\boldsymbol{v}}_1, \ldots \vec{\boldsymbol{v}}_k, \vec{\boldsymbol{v}}_{k+1}, \ldots \vec{\boldsymbol{v}}_n\}$ are linearly independent, and $\beta = \beta_1 \cup \beta_2$, with $\mathcal{V} = \operatorname{Ker} S \oplus \operatorname{Im} S$. □

For ease, suppose we call $\mathcal{W}_1 = K_\lambda(T)$ and $\mathcal{W}_2 = \operatorname{Im} S$. We know that $\mathcal{W}_1 = K_\lambda(T)$ is T-invariant, but we now need to show that $\ll \ll W_2 = \operatorname{Im} S$ is also T-invariant. This would allow us to get the Jordan canonical form, as the matrix of transformation under this basis will be split into two separate blocks.

We can easily show that $\mathcal{W}_2$ is T-invariant; if $\vec{v} \in S(\vec{w})$, then $T(\vec{v}) = T(S(\vec{w})) = S(T(\vec{w}) \in S(\vec{w})$, because S commutes with T.

Theorem 27.3: Kernel and Image of Generalized Eigenspace

If we have a linear operator $T : \mathcal{V} \to \mathcal{V}$ with eigenvalue λ, then $\exists N \geq 1$ such that $K_\lambda(T) = \operatorname{Ker}(T - \lambda\mathcal{I})^N$.

Further, if $S = (T - \lambda\mathcal{I})^n$ then $\mathcal{W}_1 = K_\lambda(T) = \operatorname{Ker} S$.

If we set $\mathcal{W}_2 = \operatorname{Im} S$, then both $\mathcal{W}_1$ and $\mathcal{W}_2$ are T-invariant subspaces of $\mathcal{V}$. We also have $\mathcal{V} = \mathcal{W}_1 \oplus \mathcal{W}_2 = \operatorname{Ker} S \oplus \operatorname{Im} S$.

This means that for bases β_1 of $\operatorname{Ker} S$ and β_2 of $\operatorname{Im} S$, $\beta = \beta_1 \cup \beta_2$ is a basis of $\mathcal{V}$ and

$$[T]_\beta = \begin{bmatrix} [T_{\operatorname{Ker} S}]_{\beta_1} & 0 \\ 0 & [T_{\operatorname{Im} S}]_{\beta_2} \end{bmatrix} = \begin{bmatrix} [T_{K_\lambda(T)}]_{\beta_1} & 0 \\ 0 & [T_{\operatorname{Im} S}]_{\beta_2} \end{bmatrix} = \begin{bmatrix} [T_{\mathcal{W}_1}]_{\beta_1} & 0 \\ 0 & [T_{\mathcal{W}_2}]_{\beta_2} \end{bmatrix}.$$

We know that $T_{\mathcal{W}_1}$ has only one eigenvalue λ, and we claim that $T_{\mathcal{W}_2}$ does not have eigenvalue λ.

Why? We know that $E_\lambda \subseteq K_\lambda(T) = \mathcal{W}_2$, and $\mathcal{W}_1 \cap \mathcal{W}_2 = \{\vec{0}\}$. This means that $\mathcal{W}_2$ cannot possibly have any vectors corresponding to eigenvalue λ.

Lemma 27.4

Let $T : \mathcal{V} \to \mathcal{V}$ be a linear operator, where $\operatorname{char}_T(\lambda)$ splits over F, and $\dim \mathcal{V} = n$.

Then, $\mathcal{V}$ breaks as a direct sum of T's generalized eigenspaces:

$$\mathcal{V} = K_{\lambda_1} \oplus K_{\lambda_2}(T) \oplus \cdots \oplus K_{\lambda_k}(T),$$

where $\lambda_1, \dots, \lambda_k$ are distinct.

That is, for any bases $\{\beta_j \text{of } K_{\lambda_j}(T)\}_{j=1}^k$, where $\beta = \cup \beta_i$ is a basis of $\mathcal{V}$.

Further, $[T]_\beta$ splits as blocks along the diagonal, where each block $T_{K_{\lambda_j}(T)}$ is the restriction of T onto each general eigenspace, which has a unique eigenvalue λ_j.

Proof. We proceed by strong induction on $\dim \mathcal{V}$.

With the base case, we have $\dim \mathcal{V} = 1$, and $T(\vec{v}) = \lambda_1 \vec{v}$. We have $[T]_\beta = (\lambda_1)_{1\times 1}$, and $\mathcal{V} = K_{\lambda_1}(T)$.

Let us assume that the claim is true for any $\dim \mathcal{V} \leq n$ for some $n \geq 1$.

Let $\dim \mathcal{V} = n + 1$. Since $\operatorname{char}_T(\lambda)$ splits over F, let λ_1 be one of its roots; λ_1 is an eigenvalue of T.

We established earlier that $\mathcal{V} = K_{\lambda_1}(T) \oplus \mathcal{W}$, or $\mathcal{V} = \operatorname{Ker} S \oplus \operatorname{Im} S$. We know that $\mathcal{W}$ is T-invariant, and that λ_1 is not an eigenvalue of $T_\mathcal{W}$.

We know that

$$\operatorname{char}_T(\lambda) = \operatorname{char}_{T_{\mathcal{W}_1}} \cdot \operatorname{char}_{T_\mathcal{W}}(\lambda) = (\lambda - \lambda_1)^{m_1} \cdot g(\lambda).$$

We must have that $m_1 = \dim \mathcal{W}_1$, and as such $\dim \mathcal{W} \leq n$, because $g(\lambda)$ has degree strictly less than $\operatorname{char}_T(\lambda)$.

By the inductive hypothesis for $T_\mathcal{W}$, we can split $\mathcal{W} = K_{\lambda_2}(T_\mathcal{W}) \oplus \cdots \oplus K_{\lambda_k}(T_\mathcal{W})$. This means that

$$\mathcal{V} = K_{\lambda_1}(T) \oplus K_{\lambda_2}(T_\mathcal{W}) \oplus \cdots \oplus K_{\lambda_k}(T_\mathcal{W}).$$

It still remains to show that $K_{\lambda_i}(T_\mathcal{W}) = K_{\lambda_i}(T)$, for all $i = 2, \dots, k$, and is left as an exercise. □

11/1/2021

Lecture 28

Constructing Jordan Bases

Recall that last time we established how the vector space $\mathcal{V}$ can be split as direct sums via $K_{\lambda_i}(T)$.

Today, we will only focus on the case where T has only one eigenvalue λ_0 on the whole space $\mathcal{V}$; that is, $\mathcal{V} = K_{\lambda_0}(T)$.

The characteristic polynomial in this case is $\mathrm{char}_T(\lambda) = \pm(\lambda - \lambda_0)^n$ (where $\dim \mathcal{V} = n$).

Example 28.1

$$[T]_e = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 3 & -1 \\ 0 & 1 & 1 \end{bmatrix}.$$

We have the characteristic polynomial $\mathrm{char}_T(\lambda) = -(\lambda-2)(\lambda^2 - 4\lambda + 4) = -(\lambda-2)^3$.

Hence, we have only one eigenvalue $\lambda_0 = 2$. This means that we must look for γ such that $[T]_\gamma = \begin{bmatrix} 2 & * & 0 \\ 0 & 2 & * \\ 0 & 0 & 2 \end{bmatrix}$, where the asterisks can either be 0 or 1.

To brainstorm, suppose $[T]_\gamma = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$. What do we know about this Jordan basis? We must have

$$\begin{aligned} T(\vec{\boldsymbol{v}}_3) &= 2\vec{\boldsymbol{v}}_3 + \vec{\boldsymbol{v}}_2 \\ T(\vec{\boldsymbol{v}}_2) &= 2\vec{\boldsymbol{v}}_2 + \vec{\boldsymbol{v}}_1 \\ T(\vec{\boldsymbol{v}}_1) &= 2\vec{\boldsymbol{v}}_1 \end{aligned}$$

Equivalently, if we have $U = T - \lambda_0\mathcal{I} = T - 2\mathcal{I}$. We then have

$$\begin{aligned} U(\vec{\boldsymbol{v}}_3) &= \vec{\boldsymbol{v}}_2 \\ U(\vec{\boldsymbol{v}}_2) &= \vec{\boldsymbol{v}}_1 \\ U(\vec{\boldsymbol{v}}_1) &= \vec{\boldsymbol{0}} \end{aligned}$$

Here, we have $[U]_\gamma = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. We like this because it's simpler. What are we doing here?

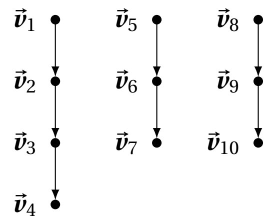
$$\vec{\boldsymbol{v}}_3 \xrightarrow{U} \vec{\boldsymbol{v}}_2 \xrightarrow{U} \vec{\boldsymbol{v}}_1 \xrightarrow{U} \vec{\boldsymbol{0}}.$$

Consecutively applying U gives us the next generalized eigenvector. We call this chain a J-cycle.

Hypothetically, suppose we have

$$[T]_\gamma = \begin{bmatrix} J_4(2) & & 0 \\ & J_3(2) & \\ 0 & & J_3(2) \end{bmatrix}.$$

Each one of these blocks has their own chain; in this case, we have a 10×10 matrix, and $\vec{\boldsymbol{v}}_4$, $\vec{\boldsymbol{v}}_7$, and $\vec{\boldsymbol{v}}_{11}$ will be our generating vectors for these three chains. We can replace the vectors by dots (and omitting the zero vector, as it is always there) to give us a dot diagram.



Again, let us look at what is happening here. Each one of these chains (U-cycles) give us a basis $\{\vec{v}, U\vec{v}, U^2\vec{v}, U^3\vec{v}\}$ or $\{\vec{v}, U\vec{v}, U^2\vec{v}\}$ depending on the chain. (Note that in this case, the stabilizing constant is $N = 4$, as this is the length of the longest chain).

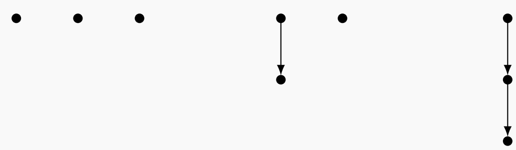
Example 28.2

If we go back to our earlier example, recall that we might have a few possibilities:

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \quad \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \quad \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}.$$

(Note that the case where we have a $0, 1$ above the diagonal, the matrix is just a reordering of the Jordan blocks.)

Which is the true Jordan form? For each of these situations, there are three different dot diagrams. We'd have



We can see that the first row (the ends of the chains) is distinct between these possibilities; the number of dots corresponds to the number of eigenvectors, i.e. the number of dots is equal to the dimension of the eigenspace.

Hence, we need to find $E_2(T)$.

$$[T - 2\mathcal{I}]_2 = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & -1 \end{bmatrix} \implies \begin{bmatrix} 0 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \implies \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \implies x_2 = x_3 = 0.$$

This means that $\operatorname{Ker}(T - 2\mathcal{I}) = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$. Since we have only one (independent) eigenvector, the correct dot diagram is the last; we have a chain of three vectors.

If the Jordan canonical form of this matrix exists, then we have

$$T \sim J_3(2) = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}.$$

How do we find the basis? We already have the eigenvector $\vec{v}_1$ at the top of the chain. We need to calculate $\vec{v}_2$ such that $U\vec{v}_2 = \vec{v}_1$. This is a non-homogeneous system, equivalent to $(T - 2\mathcal{I})\vec{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$.

If we were to solve this system, we have $x_2 = -1$ and $x_3 = -1$, meaning $\vec{\boldsymbol{v}}_2 = \left\{ \begin{bmatrix} t \\ -1 \\ -1 \end{bmatrix} \mid t \in F \right\}$. We can choose $t = 0$ to get $\vec{\boldsymbol{v}}_2 = \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix}$.

Knowing $\vec{\boldsymbol{v}}_2$, we still need to look for $\vec{\boldsymbol{v}}_3$ such that $U\vec{\boldsymbol{v}}_3 = \vec{\boldsymbol{v}}_2$. Again, we have to solve the system $(T - 2\mathcal{I})\vec{\boldsymbol{v}}_3 = \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix}$.

If we were to solve this system, we have $x_2 = 0$ and $x_3 = 1$, meaning $\vec{\boldsymbol{v}}_3 = \left\{ \begin{bmatrix} t \\ 0 \\ 1 \end{bmatrix} \mid t \in F \right\}$. We can choose $t = 0$ to get $\vec{\boldsymbol{v}}3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

We're now done! We have the basis

$$\gamma = \left\{ \vec{\boldsymbol{v}}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \vec{\boldsymbol{v}}_2 = \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix}, \vec{\boldsymbol{v}}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

These vectors are linearly independent, and form a J-cycle for $\lambda = 2$. Specifically, $[T]_\gamma = J_3(2)$, as desired.

28.1 Finding the JCF and a Jordan basis

Suppose we have a linear operator $T : \mathcal{V} \to \mathcal{V}$, where $\dim \mathcal{V} = n$, over a field F.

1. Split over F: $\operatorname{char}_T(\lambda) = \pm(\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2} \cdots (\lambda - \lambda_k)^{n_k}$, where $\lambda_i \in F$.
2. Stabilizing constants: for each λ_i, we want to find ℓ_i; i.e. set $U = T - \lambda_i \mathcal{I}$, and find the ℓ_i where $\operatorname{rank} U^{\ell_i} = \operatorname{rank} U^{\ell_i + 1}$.
3. Basis for $K_{\lambda_i}(T)$: for each λ_i, find a basis for $\operatorname{Ker}(T - \lambda_i \mathcal{I}) = E_{\lambda_i}(T)$.

   Suppose we had a basis $\{\vec{\boldsymbol{v}}_1^{(1)}, \vec{\boldsymbol{v}}_1^{(2)}, \ldots\}$. Now we know the number of chains in the dot diagram. The dimension of this eigenspace tells us how many Jordan blocks we have corresponding to this eigenvalue λ_i.

   For each of the basis vectors we found, we want to solve the system $(T - \lambda_i \mathcal{I})\vec{\boldsymbol{x}} = \vec{\boldsymbol{v}}_1^{(j)}$ for each basis vector we found, working our way up through the recursion for each eigenvector until we hit $\vec{\boldsymbol{x}} = \vec{\boldsymbol{0}}$. This gives us a J-cycle generated by each eigenvector.

   The order of these vectors matters; we start from the eigenvector and go down the chain (backwards). Rearranging will mess up the Jordan form, as the blocks no longer have a nice diagonal form.

The issue here is that each submatrix is just some random basis (it depends on our choice of t's!); next time, we will continue and find the Jordan form of the submatrices.

11/5/2021

Lecture 29

Constructing Jordan Bases II

11/8/2021

Lecture 30

JCF VI and Applications

Let us summarize an approach to compute the JCF and the Jordan basis:

Going top-down, we add e_j new vectors in a new row, and this gets us the JCF.

Going bottom-up, we pick r_j new vectors and U-propagate them, and this process gets us the Jordan basis.

Here, we have

$$\begin{aligned} \ell &= \text{stabilizing constant} \\ d_j &= \dim \operatorname{Ker} U^j && j = 1, \ldots, \ell \\ e_j &= d_j - d_{j-1} \\ r_j &= e_j - e_{j+1} = 2d_j - d_{j-1} - d_{j+1} \end{aligned}$$

In this Jordan basis γ, T is in JCF:

$$[T]_\gamma = \mathbf{J}, \qquad \mathbf{A} = \mathbf{QJQ}^{-1}.$$

Further, there is a bijection between the set of all Jordan chains and all Jordan blocks. Each Jordan block corresponds to a T-cyclic space generated by the bottom/last vector in the Jordan chain. (We arrange the vectors top-down when forming the Jordan basis.)

When does the simplified algorithm only works in extreme cases: when $\dim E_{\lambda_i} = 1$ or $\dim E_{\lambda_i} = n_i$. That is, the simplified algorithm only works if we have one chain or if we have one row.

30.1 Cayley-Hamilton Theorem

Recall that Cayley-Hamilton says that if $\mathbf{A} \in \mathcal{M}_{n\times n}(F)$ where $F = \mathbb{Q}, \mathbb{R}, \mathcal{C}$, then $\operatorname{char}_{\mathbf{A}}(\mathbf{A}) = \mathbf{0}_{n\times n}$.

Let $\operatorname{JCF}(\mathbf{A}) = \mathbf{J}$. Recall that if Cayley-Hamilton works for $\mathbf{A}$, then CHT works for $\mathbf{J}$, and vice versa.

Our next simplification involves the Jordan blocks; if $\mathbf{B}$ is block-diagonal, then the kth power $\mathbf{B}^k$ (for $k \geq 1$) will still be a block diagonal matrix:

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_1 & & \\ & \mathbf{B}_2 & \\ & & Bmat_3 \end{bmatrix} \implies \mathbf{B}^k = \begin{bmatrix} \mathbf{B}_1^k & & \\ & \mathbf{B}_2^k & \\ & & \mathbf{B}_3^k \end{bmatrix}.$$

However, if $f(t)$ is a polynomial, then we also have

$$f(\mathbf{B}) = \begin{bmatrix} f(\mathbf{B}_1) & & \\ & f(\mathbf{B}_2) & \\ & & f(\mathbf{B}_3) \end{bmatrix}.$$

This means that all we need to concentrate on is what happens for each block.

The characteristic polynomial $\operatorname{char}_{\mathbf{A}}(t) = f(t)$ is a polynomial, and we want $f(\mathbf{J}) = \mathbf{0}$. This happens if and only if each Jordan block also becomes zero. That is, $f(\mathbf{J}_i) = \mathbf{0}$ for Jordan blocks $\mathbf{J}_i$.

However, notice that if we plug in a Jordan block into the characteristic polynomial, we have something of the form $(\mathbf{J}_i - \lambda_i \mathbf{I})^k$, where k is the algebraic multiplicity of λ_i.

Notice that this is just the super-halloween operator; that is, $(\mathbf{J}_i - \lambda_i \mathbf{I})^k$ will always be a zero matrix, because we have zeroes along the diagonal, and the exponent will always propagate the zeroes throughout the rest of the matrix.

It turns out that Cayley-Hamilton also works with other fields, like $\mathbb{Z}_2, \mathbb{Z}_3, \dots, \mathbb{Z}_p$. As long as there exists a larger field where all polynomials split, then Cayley-Hamilton will work.

For example, if we're working over $\mathbb{Q}$, then we can extend all the way to $\mathbb{C}$, and all polynomials in $\mathbb{C}$ will split ($\mathbb{C}$ turns out to be the "algebraic closure of $\mathbb{R}$").

If we were working over $\mathbb{Z}_3$, then we can extend to $\overline{\mathbb{Z}}_3$, the "closure of $\mathbb{Z}_3$", where all polynomials split.

30.2 Uniqueness of JCF

We can show that any matrix has a unique Jordan canonical form (up to rearrangement of Jordan blocks).

If $\mathbf{A}$ has two JCF's, then $\mathbf{A} \sim \mathbf{J}_1$ and $\mathbf{A} \sim \mathbf{J}_2$, and as such $\mathbf{J}_1 \sim \mathbf{J}_2$.

We know that $\text{char}_{\mathbf{J}_1}(t) = \text{char}_{\mathbf{J}_2}(t)$, because the two matrices are similar. We further have the same sequences of λ_i's, with the same $m_a(\lambda_i)$ for all i.

Moreover, recall that these matrices are derived from the same linear operator—as such, invariants for the linear operator will also be invariants for $\mathbf{J}_1$ and $\mathbf{J}_2$. That is, d_j, e_j, and r_j as we talked bout before.

This means that we'd come up with the same dot diagram for both matrices. However, if the dot diagrams are the same, then the two Jordan matrices $\mathbf{J}_1$ and $\mathbf{J}_2$ must be the same, up to rearrangement of the Jordan blocks.

30.3 Existence of JCF

Note that the Jordan canonical form may not necessarily exists for matrices over $\mathbb{R}$ or $\mathbb{Q}$, because the characteristic polynomial may not split over $\mathbb{R}$ or $\mathbb{Q}$.

For example, $\mathbf{A} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, which represents a 90° rotation in $\mathbb{R}^2$. We have $\text{char}_{\mathbf{A}}(t) = t^2 + 1$, which means that there is no JCF over $\mathbb{R}$.

However, we have that $\text{char}_{\mathbf{A}}(\mathbf{A}) = \mathbf{0}$ still, i.e. Cayley-Hamilton still works here.

Again, this is because the characteristic polynomial will always split over $\mathbb{C}$, so Cayley-Hamilton will still work over $\mathbb{R}$.

30.4 Applications to Recurrences

Suppose we have $a_{n+1} = 4a_n - 4a_{n-1}$, where $a_0 = a$, $a_1 = b$.

If we calculate $\text{char}_{\mathbf{A}}(\lambda)$, we get $(\lambda - 2)^2$, which gives us $\lambda_1 = \lambda_2 = 2$. This tells us that $\mathbf{A}$ may not be diagonalizable; here, $\mathbf{A}$ is

$$\begin{bmatrix} 0 & 1 \\ -4 & 4 \end{bmatrix}.$$

This is because

$$\mathbf{A} - 2\mathbf{I} = \begin{bmatrix} -2 & 1 \\ -4 & 2 \end{bmatrix} \implies \vec{\boldsymbol{x}} \in \text{span}\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}.$$

This means that our earlier method fails.

However, we can compute the JCF of $\mathbf{A}$, such that $\mathbf{A} = \mathbf{QJQ}^{-1}$ where $\mathbf{Q}$ is the Jordan basis, and $\mathbf{J}$ is the Jordan canonical form.

We want to see if we can still compute the nth power efficiently; we have $\mathbf{A}^n = \mathbf{QJ}^n\mathbf{Q}^{-1}$. However, we need to investigate what happens to the entries above the diagonal in the Jordan canonical form. We have

$$\mathbf{J}^n = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}^n = \begin{bmatrix} \lambda^n & n\lambda^{n-1} \\ 0 & \lambda^n \end{bmatrix}.$$

Here, we only have two possible entries, so we have some $x \cdot \lambda^n + y \cdot n\lambda^{n-1}$.

It turns out that in general, we always have some polynomial $f(n)$ multiplied by λ^n in the final closed-form solution.

11/10/2021

Lecture 31

Inner Products

31.1 Dot Product

Suppose we look at $\mathbb{R}^n$. In this class, we denote the dot product as a $\circ$.

Taking $\vec{v} = (a_1, a_2, \ldots, a_n)$ and $\vec{w} = (b_1, b_2, \ldots, b_n)$, we have

$$\vec{v} \circ \vec{w} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n \in \mathbb{R}.$$

Similarly,

$$\vec{v} \circ \vec{v} = a_1^2 + a_2^2 + \cdots + a_n^2 \geq 0.$$

Since $\vec{v} \circ \vec{v}$ is nonnegative, we have

$$\sqrt{\vec{v} \circ \vec{v}} = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2} \geq 0.$$

We denote this quantity as the *norm* of $\vec{v}$, i.e.

$$\|\vec{v}\| = \sqrt{\vec{v} \circ \vec{v}}.$$

Some properties:

- (Distributivity) $(\vec{v}_1 + \vec{v}_2) \circ \vec{w} = \vec{v}_1 \circ \vec{w} + \vec{v}_2 \circ \vec{w}$
- (Commutativity) $\vec{v} \circ \vec{w} = \vec{w} \circ \vec{v}$
- $\vec{v} \circ \vec{v} \geq 0$, with equality if and only if $\vec{v} = \vec{0}$

31.2 Inner product over $\mathbb{R}$

We can now extend this to vector spaces, beyond just $\mathbb{R}^n$.

Definition 31.1: Inner Product over $\mathbb{R}$

For a vector space $\mathcal{V}$ over $\mathbb{R}$, we define an *inner product* on $\mathcal{V}$ as a binary function (it takes two inputs):

$$\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \to \mathbb{R}.$$

That is,

$$\langle \vec{v}, \vec{w} \rangle \mapsto \vec{v} \circ \vec{w}.$$

Such that

- $\langle \cdot, \cdot \rangle$ is linear in the first variable:

$$\langle \vec{v}_1 + \vec{v}_2, \vec{w} \rangle = \langle \vec{v}_1, \vec{w} \rangle + \langle \vec{v}_2, \vec{w} \rangle$$
$$\langle c\vec{v}_1, \vec{w} \rangle = c\langle \vec{v}_1, \vec{w} \rangle$$

- $\langle \cdot, \cdot \rangle$ is symmetric: $\langle \vec{v}, \vec{w} \rangle = \langle \vec{w}, \vec{v} \rangle$
- $\langle \vec{v}, \vec{v} \rangle \geq 0$ with equality if and only if $\vec{v} = \vec{0}$

Example 31.2: Standard inner product on $\mathbb{R}^n$

Suppose we have two vectors:

$$\vec{v} = a_1 \vec{e}_1 + a_2 \vec{e}_2 + \cdots + a_j \vec{e}_n$$
$$\vec{w} = b_1 \vec{e}_1 + b_2 \vec{e}_2 + \cdots + b_n \vec{e}_n$$

We have
$$\langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}} \rangle = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n.$$
If instead $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ are written in the basis $\beta = \{\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2, \ldots, \vec{\boldsymbol{v}}_n\}$ for $\mathbb{R}^n$, we have the same exact definition (but with different numbers than before!):
$$\langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}} \rangle = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n,$$
with all of the previous definitions.

We call this inner product the standard inner product with respect to the basis β.

Example 31.3

Suppose we have $\mathcal{V} = \{f : [a,b] \xrightarrow{\text{cont.}} \mathbb{R}\}$. We can agree that $\mathcal{V}$ is a vector space over $\mathbb{R}$.

We define the inner product
$$\langle f, g \rangle := \int_a^b f(x) \cdot g(x) \, \mathrm{d}x.$$
We can check that this is indeed linear in the first variable:
$$\langle c_1 f_1 + c_2 f_2, g \rangle = \int_a^b (c_1 f_1 + c_2 f_2) g \, \mathrm{d}x = c_1 \int_a^b f_1 g \, \mathrm{d}x + c_2 \int_a^b f_2 g \, \mathrm{d}x = c_1 \langle f1, g \rangle + c2 \langle f2, g \rangle.$$
We can see that $\langle f, g \rangle = \langle g, f \rangle$ because the product in the integral is commutative.

Further, we see that $\langle f, f \rangle = \int_a^b f^2 \, \mathrm{d}x \geq 0$, with equality if and only if $f(x) = 0$ on $[a,b]$.

Why? Because we're looking only at continuous functions.

Here are some other properties that follow from the definition of the standard inner product on $\mathbb{R}$.

- The inner product on $\mathcal{V}_{\mathbb{R}}$ is linear in the second variable.
$$\langle \vec{\boldsymbol{v}}, c_1 \vec{\boldsymbol{w}}_1 + c_2 \vec{\boldsymbol{w}}_2 \rangle = \langle c_1 \vec{\boldsymbol{w}}_1 + c_2 \vec{\boldsymbol{w}}_2, \vec{\boldsymbol{v}} \rangle = c_1 \langle \vec{\boldsymbol{w}}_1, \vec{\boldsymbol{v}} \rangle + c_2 \langle \vec{\boldsymbol{w}}_2, \vec{\boldsymbol{v}} \rangle = c_1 \langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}}_1 \rangle + c_2 \langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}}_2 \rangle.$$
- With the norm of $\vec{\boldsymbol{v}}$ as $\|\vec{\boldsymbol{v}}\| = \sqrt{\langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{v}} \rangle}$, we have $\|c\vec{\boldsymbol{v}}\| = |c| \cdot \|\vec{\boldsymbol{v}}\|$, for all $c \in \mathbb{R}$.

  This follows because $\langle c\vec{\boldsymbol{v}}, c\vec{\boldsymbol{v}} \rangle = c^2 \langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{v}} \rangle$, and taking the square root, we have our desired result.
- $\|\vec{\boldsymbol{v}}\| = 0$ if and only if $\vec{\boldsymbol{v}} = \vec{\boldsymbol{0}}$.

  This follows from the dot product; $\langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{v}} \rangle = 0$ if and only if $\vec{\boldsymbol{v}} = \vec{\boldsymbol{0}}$.

31.3 Inner Product over $\mathbb{C}$

Firstly, some background on $\mathbb{C}$. If $z = a + bi$, we have

- $z + \bar{z} = 2a \in \mathbb{R}$
- $z \cdot \bar{z} = (a + bi)(a - bi) = a^2 + b^2 \in \mathbb{R}^+$
- $\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2$
- $\overline{z_1 \cdot z_2} = \bar{z}_1 \cdot \bar{z}_2$

Further, $z = a + bi$, where $a = \Re(z)$, and $b = \Im(z)$.

Example 31.4

In $\mathbb{C}^2$, suppose we take the toy example of

$$\vec{\boldsymbol{v}} \circ \vec{\boldsymbol{w}} = a_1\bar{a}_2 + b_1\bar{b}_2.$$

Here, we have $\vec{\boldsymbol{v}} = \langle a_1 b_1 \rangle$ and $\vec{\boldsymbol{w}} = \langle a_2, b_2 \rangle$.

This means that

$$\|\vec{\boldsymbol{v}}\| = \sqrt{\vec{\boldsymbol{v}}, \vec{\boldsymbol{v}}} = \sqrt{a_1 \cdot \bar{a}_1 + b_1 \cdot \bar{b}_1}.$$

The quantities under the square root are both real positive numbers, so we're fine.

Definition 31.5: Inner Product over $\mathbb{C}$

For a vector space $\mathcal{V}$ over $\mathbb{C}$, we define an *inner product* on $\mathcal{V}$ as a binary function (it takes two inputs):

$$\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \to \mathbb{C}.$$

That is,

$$\langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}} \rangle \mapsto \vec{\boldsymbol{v}} \circ \vec{\boldsymbol{w}}.$$

Such that

- $\langle \cdot, \cdot \rangle$ is linear in the first variable:

$$\langle \vec{\boldsymbol{v}}_1 + \vec{\boldsymbol{v}}_2, \vec{\boldsymbol{w}} \rangle = \langle \vec{\boldsymbol{v}}_1, \vec{\boldsymbol{w}} \rangle + \langle \vec{\boldsymbol{v}}_2, \vec{\boldsymbol{w}} \rangle$$
$$\langle c\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{w}} \rangle = c\langle \vec{\boldsymbol{v}}_1, \vec{\boldsymbol{w}} \rangle$$

- $\langle \cdot, \cdot \rangle$ is *hermetian*: $\langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}} \rangle = \overline{\langle \vec{\boldsymbol{w}}, \vec{\boldsymbol{v}} \rangle}$
- $\langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{v}} \rangle \geq 0$ with equality if and only if $\vec{\boldsymbol{v}} = \vec{\boldsymbol{0}}$

The properties over $\mathbb{R}$ pretty much all extend to properties over $\mathbb{C}$. However, $\mathbb{C}$ is not quite linear in the second variable:

$$\begin{aligned}\langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}}_1 + \vec{\boldsymbol{w}}_2 \rangle &= \overline{\langle \vec{\boldsymbol{w}}_1 + \vec{\boldsymbol{w}}_2, \vec{\boldsymbol{v}} \rangle} \\ &= \overline{\langle \vec{\boldsymbol{w}}_1, \vec{\boldsymbol{v}} \rangle} + \overline{\langle \vec{\boldsymbol{w}}_1, \vec{\boldsymbol{v}} \rangle} \\ &= \langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}}_1 \rangle + \langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}}_1 \rangle\end{aligned}$$

This is fine, but when we add a constant, we have something else:

$$\begin{aligned}\langle \vec{\boldsymbol{v}}, c\vec{\boldsymbol{w}} \rangle &= \overline{\langle c\vec{\boldsymbol{w}}, \vec{\boldsymbol{v}} \rangle} \\ &= \bar{c}\overline{\langle c\vec{\boldsymbol{w}}, \vec{\boldsymbol{v}} \rangle} \\ &= \bar{c}\langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}} \rangle\end{aligned}$$

Example 31.6

Suppose we take a basis β for $\mathcal{V}$ over $\mathbb{C}$.

We can see that $\langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}} \rangle = 2a_1\bar{b}_1 + 2a_2\bar{b}_2 + \cdots + 2a_n\bar{b}_n$ is also a valid inner product; all the properties hold.

Example 31.7

With $\mathcal{V}$ as $\mathbb{R}^n$ or $\mathbb{C}^n$, we should be able to see that

$$z \cdot \bar{z} = a^2 + b^2 = |z|^2.$$

This means that

$$\|\vec{\boldsymbol{v}}\| = \|(a_1, a_2, \dots, a_n)_e\| = \sqrt{a_1 \cdot \bar{a}_1 \cdots + a_n \cdot \bar{a}_n} = \sqrt{|a_1|^2 + \cdots + |a_n|^2}.$$

This is essentially a generalized Pythagorean Theorem!

Note that our definition for $\mathbb{V}_{\mathbb{C}}$ does not work for $\mathcal{V}_{\mathbb{C}}$. If we take $\|(0, i)\|$, we get $\sqrt{0 \cdot 0 + i \cdot i} = \sqrt{-1} = i \notin \mathbb{R}$.

However, our definition for $\mathcal{V}_{\mathbb{C}}$ *does* also work for $\mathcal{V}_{\mathbb{R}}$; the reals are also complex numbers.

31.4 Cauchy-Bunyakovski-Schwartz Inequality

Theorem 31.8: CBS Ineqeuality

Suppose we have a vector space $\mathcal{V}_{\mathbb{R}}$ or $\mathcal{V}_{\mathbb{C}}$ with inner product $\langle \cdot, \cdot \rangle$, and vectors $\vec{\boldsymbol{v}}, \vec{\boldsymbol{w}} \in \mathcal{V}$.

We then have

$$|\langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}} \rangle| \leq \|\vec{\boldsymbol{v}}\| \cdot \|\vec{\boldsymbol{w}}\|.$$

Further, we have equality if and only if $\vec{\boldsymbol{v}}$ and $\vec{\boldsymbol{w}}$ are linearly dependent.

Proof. We start with an observation. Suppose we have $c \in \mathbb{C}$; we then have

$$\begin{aligned}
0 \leq \|\vec{\boldsymbol{v}} + c\vec{\boldsymbol{w}}\|^2 &= \langle \vec{\boldsymbol{v}} + c\vec{\boldsymbol{w}}, \vec{\boldsymbol{v}} + c\vec{\boldsymbol{w}} \rangle \\
&= \langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{v}} \rangle + \bar{c}\langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}} \rangle + c\langle \vec{\boldsymbol{w}}, \vec{\boldsymbol{v}} \rangle + c\bar{c}\langle \vec{\boldsymbol{w}}, \vec{\boldsymbol{w}} \rangle \\
&= \|\vec{\boldsymbol{v}}\|^2 + \bar{c}\langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}} \rangle + c\overline{\langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}} \rangle} + c\bar{c}\|\vec{\boldsymbol{w}}\|^2
\end{aligned}$$

We would like to simplify this as much as possible. We would like to eliminate the last two terms; choosing

$$\bar{c} = -\frac{\overline{\langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}} \rangle}}{\|\vec{\boldsymbol{w}}\|^2},$$

we can eliminate all of the last two terms. Substituting the $\bar{c}$ in the second term, we have

$$\begin{aligned}
0 &\leq \|\vec{\boldsymbol{v}}\|^2 - \frac{\overline{\langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}} \rangle}}{\|\vec{\boldsymbol{w}}\|^2} \cdot \langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}} \rangle \\
0 &\leq \|\vec{\boldsymbol{v}}\|^2 \|\vec{\boldsymbol{w}}\|^2 - ||\vec{\boldsymbol{v}}, \vec{\boldsymbol{w}}||^2 \\
|\langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}} \rangle|^2 &\leq \|\vec{\boldsymbol{v}}\|^2 \cdot \|\vec{\boldsymbol{w}}\|^2 \\
|\langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}} \rangle| &\leq \|\vec{\boldsymbol{v}}\| \cdot \|\vec{\boldsymbol{w}}\|
\end{aligned}$$

□

Example 31.9

Let us look at an application of this in $\mathbb{R}^2$.

If we have two vectors $\vec{v}$ and $\vec{w}$ in $\mathbb{R}^2$, and define α to be the angle between the two vectors, we have

$$\langle \vec{v}, \vec{w} \rangle = \vec{v} \circ \vec{w} = \|\vec{v}\| \cdot \|\vec{w}\| \cdot \cos(\alpha).$$

The extremes here are if $\cos(\alpha) = \pm 1$. We have $\cos(\alpha) = 1$ if $\alpha = 0$, and $\cos(\alpha) = -1$ if $\alpha = \pi$. The former case happens when $\vec{v}$ and $\vec{w}$ are parallel. The latter case happens when $\vec{v}$ and $\vec{w}$ are orthogonal.

This means that

$$-\|\vec{v}\| \cdot \|\vec{w}\| \leq \langle \vec{v}, \vec{w} \rangle \leq \|\vec{v}\| \cdot \|\vec{w}\| \implies |\langle \vec{v}, \vec{w} \rangle| \leq \|\vec{v}\| \cdot \|\vec{w}\|.$$

This leads us to question: can we define an angle between $\vec{v}$ and $\vec{w}$ in an arbitrary vector space $\mathcal{V}$?

We can; we'd define

$$\cos(\alpha) = \frac{\langle \vec{v}, \vec{w} \rangle}{\|\vec{v}\| \cdot \|\vec{w}\|}.$$

We can further verify that the RHS is always between -1 and 1, so this is a valid definition!

11/12/2021

Lecture 32

Applications of Inner Products, Orthogonal Bases

32.1 Triangle Inequality

Theorem 32.1: Triangle Inequality

Working under $\mathcal{V}_{\mathbb{C}}$ or $\mathcal{V}_{\mathbb{R}}$, we have that for all $\vec{v}, \vec{w} \in \mathcal{V}$:

$$\|\vec{v} + \vec{w}\| \leq \|\vec{v}\| + \|\vec{w}\|.$$

Equality occurs if and only if $\vec{v} \parallel \vec{w}$ over the reals. That is, when $\vec{v} = c\vec{w}$ for $c \in \mathbb{R}^+$ or $\vec{w} = \vec{0}$.

Proof. We will split this into two cases.

If we are working over $\mathbb{R}$, we have that

$$\begin{aligned}
\|\vec{v} + \vec{w}\|^2 &= \langle \vec{v} + \vec{w}, \vec{v} + \vec{w} \rangle \\
&= \|\vec{v}\|^2 + \|\vec{w}\|^2 + \langle \vec{v}, \vec{w} \rangle + \langle \vec{w}, \vec{v} \rangle \\
&= \|\vec{v}\|^2 + \|\vec{w}\|^2 + 2\langle \vec{v}, \vec{w} \rangle && (\langle \vec{v}, \vec{w} \rangle = \langle \vec{w}, \vec{v} \rangle \text{ over } \mathbb{R}) \\
&\leq \|\vec{v}\|^2 + \|\vec{w}\|^2 + 2|\langle \vec{v}, \vec{w} \rangle| \\
&\leq \|\vec{v}\|^2 + \|\vec{w}\|^2 + 2\|\vec{v}\| \cdot \|\vec{w}\| && \text{(CBS)} \\
&= (\|\vec{v}\| + \|\vec{w}\|)^2 \\
\|\vec{v} + \vec{w}\| &= \|\vec{v}\| + \|\vec{w}\|
\end{aligned}$$

Further, note that we have equality if and only if $\langle \vec{v}, \vec{w} \rangle = |\langle \vec{v}, \vec{w} \rangle|$ as well as $\vec{v} = c\vec{w}$ for some $c \in \mathbb{R}$. The former takes care of the first inequality conversion (with absolute value), and the latter takes care of the second inequality conversion (with CBS). Together, this implies that we must have

$$\langle c\vec{w}, \vec{w} \rangle = c\|\vec{w}\| = |c|\|\vec{w}\|^2.$$

This only occurs if $c \geq 0$ or $\vec{w} = \vec{0}$; this is the exact equality condition as claimed.

If we are working over $\mathbb{C}$, we need to take into account the fact that $\langle \vec{w}, \vec{v} \rangle = \overline{\langle \vec{v}, \vec{w} \rangle}$. This means that instead of $2\langle \vec{v}, \vec{w} \rangle$, we'd have $\langle \vec{v}, \vec{w} \rangle + \overline{\langle \vec{v}, \vec{w} \rangle} = 2\Re\langle \vec{v}, \vec{w} \rangle$.

Note that we can still bound this expression by $2|\langle \vec{v}, \vec{w}\rangle|$ as in the case for $\mathbb{R}$. This is because for a complex number $z = a + bi$, we have $|z| = \sqrt{a^2 + b^2}$, with $\Re z = a$. We know that $a \le \sqrt{a^2 + b^2}$; visually the LHS is the horizontal leg of the right triangle in the complex plane, and the RHS is the hypotenuse of the triangle.

This means that we can still follow the rest of the proof and arrive at the same result over $\mathbb{C}$. $\square$

Example 32.2

In the simplest non-trivial case of $\mathbb{R}^1$, we can see that $|a+b| \le |a| + |b|$, and $|a+b| = |a| + |b|$ if and only if a and b have the same signs.

32.1.1 Applications

Example 32.3

Suppose we're working under $\mathcal{V} = \mathbb{R}^n$. We have $\vec{v} = (a_1, a_2, \dots, a_n)$ and $\vec{w} = (b_1, b_2, \dots, b_n)$, with $\vec{v} + \vec{w} = (a_1 + b_1, \dots, a_n + b_n)$.

We then have by the triangle inequality

$$\|\vec{v} + \vec{w}\| \le \|\vec{v}\| + \|\vec{w}\| \implies \sqrt{(a_1+b_1)^2 + \cdots + (a_n + b_n)^2} \le \sqrt{a_1^2 + \cdots + a_n^2} + \sqrt{b_1^2 + \cdots + b_n^2}.$$

We further have equality if and only if $a_j = cb_j$ for all j and some $c \ge 0$ (or all $b_j = 0$).

32.2 Orthogonality

Besides 0 and π (i.e. parallel), are there other distinguished angles between vectors?

Under the inner product space $\mathcal{V}_F$, we have a few definitions.

Definition 32.4: Orthogonal Vectors

$\vec{v}, \vec{w} \in \mathcal{V}$ are *orthogonal* if $\langle \vec{v}, \vec{w}\rangle = 0$. (This means that $\vec{v} \perp \vec{w}$; the two are perpendicular).

Definition 32.5: Unitary Vectors

$\vec{v} \in \mathcal{V}$ is *unitary* (unit vector) if $\|\vec{v}\| = 1$, i.e. $\langle \vec{v}, \vec{v}\rangle = 1$.

We can create such a unit vector from an arbitrary nonzero vector $\vec{v} \in \mathcal{V}$ with $\vec{u} = \frac{\vec{v}}{\|\vec{v}\|}$; this makes $\|\vec{u}\| = 1$.

Definition 32.6: Orthogonal and Orthonormal Sets

Suppose $S = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\} \subset V$.

S is an *orthogonal* set of vectors if $\vec{v}_i \perp \vec{v}_j$ for all $i \ne j$.

S is an *orthonormal* set of vectors if $\vec{v}_i \perp \vec{v}_j$ for all $i \ne j$ and $\|\vec{v}_i\| = 1$ for all i. That is, S is orthonormal if it is orthogonal and contains only unit vectors.

Lemma 32.7

If $S = \{\vec{v}_1, \dots, \vec{v}_k\}$ is an orthogonal set of nonzero vectors, then S is linearly independent.

Proof. Let us look at the linear relation $a_1\vec{v}_1 + \cdots + a_k\vec{v}_k = \vec{0}$ for some $a_j \in F$. If we take the inner product with $\vec{v}_i$ and expand with linearity, we have

$$\left\langle \sum_{j=1}^{k} a_j \vec{v}_j, \vec{v}_i \right\rangle = \left\langle \vec{0}, \vec{v}_i \right\rangle = \vec{0}.$$

Recalling that S is an orthogonal set of vectors, we must have that $\left\langle \vec{v}_j, \vec{v}_i \right\rangle = 0$ for all $j \neq i$. This leaves just

$$\langle a_i \vec{v}_i, \vec{v}_i \rangle = a_i \|\vec{v}_i\|^2 = 0.$$

Since $\vec{v}_i \neq \vec{0}$, this forces $a_i = 0$. Further, we chose an arbitrary i here, so repeating the process for all i, we'd end up with the fact that $a_i = 0$ for all i, meaning S is linearly independent. □

Corollary 32.8

If $\dim \mathcal{V} = n$, and $S = \{\vec{v}_1, \dots, \vec{v}_n\}$, is an orthogonal set of nonzero vectors, then S is a basis for $\mathcal{V}$.

11/15/2021

Lecture 33

Gram-Schmidt Process

Recall that over the inner product space $\mathcal{V}_F$, $\beta = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ is an orthonormal basis if and only if

- $\left\langle \vec{v}_i, \vec{v}_j \right\rangle = \delta_{ij}$ (that is, 1 if $i = j$, and 0 if $i \neq j$)
- $\vec{v}_i \perp \vec{v}_j$ for all $i \neq j$
- $\|\vec{v}_i\| = 1$ for all i

33.1 Coordinates in an Orthonormal Basis

Let us look at how we can compute the coordinates of $\vec{v}$ with respect to the orthonormal basis $\beta = \{\vec{v}_1, \dots, \vec{v}_n\}$ of $\mathcal{V}$. That is, we would like the vector of a_i's where $\vec{v} = a_1\vec{v}_1 + a_2\vec{v}_2 + \cdots + a_n\vec{v}_n$.

Notice that since $\vec{v}_i$ is orthogonal to all other vectors in β, we have

$$\langle \vec{v}, \vec{v}_i \rangle = \left\langle \sum_{i=j}^{n} a_j \vec{v}_j, \vec{v}_i \right\rangle = \langle a_i \vec{v}_i, \vec{v}_i \rangle.$$

Further, since $\vec{v}_i$ is a unit vector (since β is orthonormal), we have

$$\langle a_i \vec{v}_i, \vec{v}_i \rangle = a_i \|\vec{v}_i\|^2 = a_i.$$

This means that we can find the coordinates really fast if we know $\langle \cdot, \cdot \rangle$ and if we are given an orthonormal basis:

$$[\vec{v}]_\beta = \begin{bmatrix} \langle \vec{v}, \vec{v}_1 \rangle \\ \langle \vec{v}, \vec{v}_2 \rangle \\ \cdots \\ \langle \vec{v}, \vec{v}_n \rangle \end{bmatrix}_\beta .$$

33.2 Gram-Schmidt Process

Theorem 33.1: Gram-Schmidt Process

In an inner product space $\mathcal{V}_F$, suppose we have a basis $\beta = \{\vec{w}_1, \dots, \vec{w}_n\}$ for $\mathcal{V}$. We claim that we can always find an orthogonal basis $\beta' = \{\vec{v}_1, \dots, \vec{v}_n\}$ for $\mathcal{V}$ such that

$$\vec{v}_i \in \operatorname{span}\left\{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_i\right\} \quad \forall i = 1, 2, \dots, n.$$

Proof. We provide a constructive proof by induction on $k = 1, \dots, n$.

As a base case, we have $\vec{v}_1 = \vec{w}_1 \in \operatorname{span}\left\{\vec{w}_1\right\}$. For the inductive hypothesis, suppose that we have found an orthogonal set of nonzero vectors $\{\vec{v}_1, \dots, \vec{v}_{k-1}\}$ that satisfy the claim for some $k \in [2, n)$.

In the inductive step, we generate $\vec{v}_k$. Let us set $\vec{v}_k = \vec{w}_k + a_1 \vec{v}_1 + a_2 \vec{v}_2 + \cdots + a_{k-1} \vec{v}_{k-1}$ for some set of $a_i \in F$; our goal is to find these a_i's.

We would like $\langle \vec{v}_k, \vec{v}_i \rangle = 0$ for all i. However, we know that

$$\langle \vec{v}_k, \vec{v}_i \rangle = \left\langle \vec{w}_k + \sum_{j=1}^{k-1} a_j \vec{v}_j, \vec{v}_i \right\rangle = \langle \vec{w}_k, \vec{v}_i \rangle + \langle a_i \vec{v}_i, \vec{v}_i \rangle.$$

As such, we have that

$$\langle \vec{w}_k, \vec{v}_i \rangle + a_i \langle \vec{v}_i, \vec{v}_i \rangle = 0 \implies -\langle \vec{w}_k, \vec{v}_i \rangle = a_i \|\vec{v}_i\|^2 \implies \boxed{a_i = -\frac{\langle \vec{w}_k, \vec{v}_i \rangle}{\|\vec{v}_i\|^2}}.$$

Putting all of this together, we have that

$$\boxed{\vec{v}_k = \vec{w}_k - \sum_{i=1}^{k-1} \frac{\langle \vec{w}_K, \vec{v}_i \rangle}{\|\vec{v}_i\|^2} \vec{v}_i}.$$

Note that must have $\vec{v}_k \neq 0$ otherwise $\vec{w}_k \in \operatorname{span}\left\{\vec{w}_1, \dots, \vec{w}_{k-1}\right\}$, meaning we did not have a basis of $\vec{w}_i$'s to begin with. □

Corollary 33.2

Any inner product space $\mathcal{V}_F$ has an orthonormal basis.

Corollary 33.3

Any orthogonal set $\{\vec{w}_1, \dots, \vec{w}_m\}$ of nonzero vectors can be extended to an orthogonal basis of $\mathcal{V}$.

One thing to note is that the Gram-Schmidt process always preserves the first m orthogonal vectors; for $k < m$, if we already have $\vec{v}_i = \vec{w}_i$ for $i < k$, then we'd get

$$\vec{v}_k = \vec{w}_k - \sum_{i=1}^{k-1} \frac{\langle \vec{w}_k, \vec{v}_i \rangle}{\|\vec{v}_i\|^2} = \vec{w}_k - \sum_{i=1}^{k-1} \frac{\cancel{\langle \vec{w}_k, \vec{w}_i \rangle}}{\|\vec{w}_i\|^2} = \vec{w}_k - \vec{0} = \vec{w}_k.$$

33.3 Projections onto Subspaces

Given $\mathcal{W}$, a subspace of $\mathcal{V}$, and a vector $\vec{v} \notin \mathcal{W}$, we want to look for a $\vec{w} \in \mathcal{W}$ such that $(\vec{v} - \vec{w}) \perp \mathcal{W}$.

To do this, we can use Gram-Schmidt to find an orthonormal basis $\gamma = \{\vec{w}_1, \dots, \vec{w}_m\}$ for $\mathcal{W}$.

Letting $\vec{w} = a_1\vec{w}_1 + a_2\vec{w}_2 + \cdots + a_m\vec{w}_m$, we would like $\vec{v} - \vec{w} = \vec{v} - \sum_{j=1}^m a_j\vec{w}_j$ to be perpendicular to $\mathcal{W}$. This occurs if and only if $\vec{v} - \vec{w}$ is perpendicular to all $\vec{w}_1, \ldots, \vec{w}_m$:

$$\langle \vec{v} - \vec{w}, \vec{w}_i\rangle = \vec{0} = \langle \vec{v}, \vec{w}_i\rangle - a_i\cancelto{1}{\langle \vec{w}_i, \vec{w}_i\rangle} \implies a_i = \langle \vec{v}, \vec{w}_i\rangle.$$

Hence, we have

Definition 33.4: Projection Onto Subspace

The projection of $\vec{v} \in \mathcal{V}$ onto the subspace $\mathcal{W}$ is defined as

$$\vec{w} = \operatorname{proj}_{\mathcal{W}}(\vec{v}) = \langle \vec{v}, \vec{w}_1\rangle\vec{w}_1 + \cdots + \langle \vec{v}, \vec{w}_m\rangle\vec{w}_m \in \mathcal{W}.$$

If we look at an orthonormal basis $\{\vec{w}_1, \ldots, \vec{w}_m\}$ for $\mathcal{W}$, we can extend it using Gram-Schmidt to an orthonormal basis $\{\vec{w}_1, \ldots, \vec{w}_m, \vec{v}_{m+1}, \ldots, \vec{v}_n\}$ for $\mathcal{V}$. Observe that for any $\vec{v} \in \mathcal{V}$:

$$\vec{v} = \underbrace{\sum_{i=1}^n \langle \vec{v}, \vec{v}_i\rangle\vec{v}_i}_{\operatorname{proj}_{\mathcal{V}}(\vec{v}) = \vec{v}} = \underbrace{\sum_{i=1}^m \langle \vec{v}, \vec{w}_i\rangle\vec{w}_i}_{\operatorname{proj}_{\mathcal{W}}(\vec{v})} + \underbrace{\sum_{j=m+1}^n \left\langle \vec{v}, \vec{v}_j\right\rangle\vec{v}_j}_{\vec{v} - \operatorname{proj}_{\mathcal{W}}(\vec{v})}.$$

That is, we can always break down any vector into its projection onto $\mathcal{W}$ (i.e. the component of $\vec{v}$ parallel to $\mathcal{W}$) and the component of $\vec{v}$ perpendicular to $\mathcal{W}$.

Here, we used the following lemma:

Lemma 33.5

If $\vec{u} \perp \mathcal{W}$ (that is, $\vec{u} \perp \vec{w}_1, \ldots, \vec{w}_m$, where the RHS is a basis for $\mathcal{W}$), then $\vec{u} \perp \vec{w}$ for all $\vec{w} \in \mathcal{W}$.

Proof. We have $\langle \vec{u}, \vec{w}\rangle = \langle \vec{u}, a_1\vec{w}_1 + \cdots + a_m\vec{w}_m\rangle = 0$, because $\langle \vec{u}, \vec{w}_i\rangle = 0$, so splitting the inner products up, we'd get a sum of zeros.

As such, it is enough to be perpendicular to a basis of a subspace in order to be perpendicular to the entire subspace. □

Definition 33.6: Distance to a Subspace

$\left\|\vec{v} - \operatorname{proj}_{\mathcal{W}}(\vec{v})\right\|$ is called the *distance* from $\vec{v}$ to $\mathcal{W}$.

One can prove that this is in fact the shortest distance from $\vec{v}$ to $\mathcal{W}$ by contradiction; it is omitted here for brevity (more because I'm lazy).

33.4 Orthogonal Complements

Definition 33.7: Orthogonal Complement

Suppose we have a subspace $\mathcal{W}$ of $\mathcal{V}$, where $\dim\mathcal{V} < \infty$. We define

$$\mathcal{W}^\perp = \{\vec{v} \in \mathcal{V} \mid \vec{v} \perp \mathcal{W}\}$$

to be the *orthogonal complement* of $\mathcal{W}$ in $\mathcal{V}$.

Here are some facts:

Lemma 33.8

$\mathcal{W}^\perp \leqq \mathcal{V}$; that is, $\mathcal{W}^\perp$ is a subspace of $\mathcal{V}$.

Proof. Suppose we have $\vec{v}_1, \vec{v}_2 \in \mathcal{W}^\perp$ and $c_1, c_2 \in F$. We would like to prove that $c_1\vec{v}_1 + c_2\vec{v}_2 \in \mathcal{W}^\perp$.

We can see that for any $\vec{w} \in \mathcal{W}$, we have

$$\langle c_1\vec{v}_1 + c_2\vec{v}_2, \vec{w}\rangle = c_1\langle \vec{v}_1, \vec{w}\rangle + c_2\langle \vec{v}_2, \vec{w}\rangle = 0,$$

because $\vec{v}_1$ and $\vec{v}_2$ are orthogonal to $\vec{w}$ by definition.

As such, $c_1\vec{v}_1 + c_2\vec{v}_2 \perp \mathcal{W}$ for all $\vec{w} \in \mathcal{W}$, and as such $c_1\vec{v}_1 + c_2\vec{v}_2 \in \mathcal{W}^\perp$ as desired. □

Lemma 33.9

$\mathcal{W} \cap \mathcal{W}^\perp = \{\vec{0}\}$.

Proof. Let $\vec{u} \in \mathcal{W} \cap \mathcal{W}^\perp$. We then must have $\vec{u} \perp \vec{u}$, which implies that $\langle \vec{u}, \vec{u}\rangle = \vec{0}$. By the definition of an inner product, we then must have that $\vec{u} = \vec{0}$. □

Lemma 33.10

$\mathcal{V} = \mathcal{W} \oplus \mathcal{W}^\perp$.

Proof. Let $\gamma = \{\vec{w}_1, \dots, \vec{w}_m\}$ be an orthonormal basis for $\mathcal{W}$.

By Gram-Schmidt, we can extend γ to an orthonormal basis for $\mathcal{V}$: $\beta = \{\vec{w}_1, \dots, \vec{w}_m, \vec{w}_{m+1}, \dots, \vec{w}_n\}$. Our claim is that $\mathcal{W}^\perp = \operatorname{span}\{\vec{w}_{m+1}, \dots \vec{w}_n\}$.

We can see that $\vec{v} = \vec{w} + \vec{y}$, where $\vec{w} = \operatorname{proj}_{\mathcal{W}}(\vec{v}) \in \mathcal{W}$ with $\vec{y} = \vec{v} - \vec{w}$. We can see that $\vec{y} \perp \mathcal{W}$, because $\vec{w}_i \perp \vec{w}_j$'s for $i \le m < j$, and as such $\vec{y} \in \mathcal{W}^\perp$.

Since this means that we can always split $\vec{v}$ into a sum of two vectors, one of which is in $\mathcal{W}$ and the other in $\mathcal{W}^\perp$, we must have that $\mathcal{V} = \mathcal{W} \oplus \mathcal{W}^\perp$. □

11/17/2021

Lecture 34

Adjoint Linear Operators

Definition 34.1: Adjoint of a Matrix

Suppose we have $\mathbf{A}_{n\times n}$ over $\mathbb{C}$. The *adjoint* of $\mathbf{A}$ is

$$\mathbf{A}^* = \overline{\mathbf{A}^T} = \overline{\mathbf{A}}^T.$$

Example 34.2

Suppose we have a matrix $\mathbf{A}$ with rows $\vec{\boldsymbol{v}}_i$. Given $\langle \vec{\boldsymbol{v}}_i, \vec{\boldsymbol{v}}_j \rangle = \delta_{ij}$ for all i, j, this means that

$$\sum_{k=1}^{n} a_{ik} \cdot \bar{a}_{kj} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}.$$

This reminds us of matrix multiplication; if we take the product of $\mathbf{A}$ and its adjoint $\mathbf{A}^*$, then we'd find that $\mathbf{A}\mathbf{A}^* = \mathbf{I}_n$. That is, rows of $\mathbf{A}$ are $\vec{\boldsymbol{v}}_i$'s, and the columns of $\mathbf{A}^*$ are $\bar{\vec{\boldsymbol{v}}}_j$'s, and the dot product gives us our δ_{ij}.

Further, we can show that $\mathbf{A}^T \cdot \bar{\mathbf{A}} = \mathbf{I}_n$ as well.

As such, we can see that the columns of $\mathbf{A}$ (i.e. rows of $\mathbf{A}^T$) are also orthonormal; such an $\mathbf{A}$ is called an orthogonal matrix, or a unitary matrix.

We can now switch from matrices to linear operators; two questions arise. Geometrically, what do adjoint linear operators do? Further, how do they relate to the original linear operators?

One initial attempt could be to look at a linear operator $T : \mathcal{V} \to \mathcal{V}$, and define the linear operator $T^* : \mathcal{V} \to \mathcal{V}$ such that the matrix of transformation $[T]_\beta = \mathbf{A}$ gets mapped to $\mathbf{A}^* = \bar{\mathbf{A}}^T = [T^*]_\beta$, under some orthonormal basis β for the inner product space $\mathcal{V}$.

However, this isn't a very good definition; we do not know (yet) if the matrices of T and T^* will still be adjoint in another orthonormal basis.

As such, we can try another attempt:

Definition 34.3: Adjoint of a Linear Operator

If we have the linear operator $T : \mathcal{V} \to \mathcal{V}$ where $\mathcal{V}$ is an inner product space, we define $T^* : \mathcal{V} \to \mathcal{V}$ to be the unique function such that

$$\langle T(\vec{\boldsymbol{x}}), \vec{\boldsymbol{y}} \rangle = \langle \vec{\boldsymbol{x}}, T^*(\vec{\boldsymbol{y}}) \rangle \quad \forall \vec{\boldsymbol{x}}, \vec{\boldsymbol{y}} \in \mathcal{V}.$$

Moreover, this T^* is a linear operator.

Lemma 34.4

The adjoint T^* of T as defined above is linear.

Proof. Firstly, we show that we can split sums. Suppose we take $\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}}_1, \vec{\boldsymbol{y}}_2 \in \mathcal{V}$. We have

$$\begin{aligned}
\langle \vec{\boldsymbol{x}}, T^*(\vec{\boldsymbol{y}}_1 + \vec{\boldsymbol{y}}_2) \rangle &= \langle T(\vec{\boldsymbol{x}}), \vec{\boldsymbol{y}}_1 + \vec{\boldsymbol{y}}_2 \rangle && \text{(def. } T^*) \\
&= \langle T(\vec{\boldsymbol{x}}), \vec{\boldsymbol{y}}_1 \rangle + \langle T(\vec{\boldsymbol{x}}), \vec{\boldsymbol{y}}_2 \rangle && \text{(split inner prod.)} \\
&= \langle \vec{\boldsymbol{x}}, T^*(\vec{\boldsymbol{y}}_1) \rangle + \langle \vec{\boldsymbol{x}}, T^*(\vec{\boldsymbol{y}}_2) \rangle && \text{(def. } T^*) \\
&= \langle \vec{\boldsymbol{x}}, T^*(\vec{\boldsymbol{y}}_1) + T^*(\vec{\boldsymbol{y}}_2) \rangle && \text{(combine inner prod.)} \\
\vec{\mathbf{0}} &= \langle \vec{\boldsymbol{x}}, T^*(\vec{\boldsymbol{y}}_1) + T^*(\vec{\boldsymbol{y}}_2) - T^*(\vec{\boldsymbol{y}}_1 + \vec{\boldsymbol{y}}_2) \rangle
\end{aligned}$$

Let us call the RHS $\vec{\boldsymbol{w}}$. Since this equation is true for all $\vec{\boldsymbol{x}}$, let us suppose $\vec{\boldsymbol{x}} = \vec{\boldsymbol{w}}$. This means that we must have $\langle \vec{\boldsymbol{w}}, \vec{\boldsymbol{w}} \rangle = \vec{\mathbf{0}}$, which forces $\vec{\boldsymbol{w}} = \vec{\mathbf{0}}$, meaning $T^*(\vec{\boldsymbol{y}}_1 + \vec{\boldsymbol{y}}_2) = T^*(\vec{\boldsymbol{y}}_1) + T^*(\vec{\boldsymbol{y}}_2)$, as desired.

Next, we show that we can take out scalar multiplication. Suppose we take $a \in \mathbb{C}$ and $\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}} \in \mathcal{V}$. We have

$$\begin{aligned}
\langle \vec{\boldsymbol{x}}, T^*(a\vec{\boldsymbol{y}}) \rangle &= \langle T(\vec{\boldsymbol{x}}), a\vec{\boldsymbol{y}} \rangle && \text{(def. } T^*) \\
&= \bar{a}\langle T(\vec{\boldsymbol{x}}), \vec{\boldsymbol{y}} \rangle \\
&= \bar{a}\langle \vec{\boldsymbol{x}}, T^*(\vec{\boldsymbol{y}}) \rangle
\end{aligned}$$

$$= \langle \vec{x}, aT^*(\vec{y}) \rangle \qquad (\overline{\overline{a}} = a)$$
$$\langle \vec{x}, T^*(a\vec{y}) - aT^*(\vec{y}) \rangle = 0$$

As such, by a similar proof as before, we have that $T^*(a\vec{y}) = aT^*(\vec{y})$, as desired. □

We've used the following lemma twice in the previous proof:

Lemma 34.5

If $\langle \vec{x}, \vec{w} \rangle = \langle \vec{x}, \vec{w}' \rangle$ for all $\vec{x} \in \mathcal{V}$ where $\mathcal{V}$ is an inner product space, then $\vec{w} = \vec{w}'$.

Proof. If $\langle \vec{x}, \vec{w} - \vec{w}' \rangle = 0$ for all $\vec{x} \in \mathcal{V}$, then it holds for $\vec{x} = \vec{w} - \vec{w}'$ in particular as well. This means that $\langle \vec{w} - \vec{w}', \vec{w} - \vec{w}' \rangle = 0$, meaning $\vec{w} - \vec{w}' = \vec{0}$ and $\vec{w} = \vec{w}'$. □

Lemma 34.6

Under any basis β for the inner product space $\mathcal{V}$, the adjoint T^* of T has a matrix of transformation $[T^*]_\beta = \overline{\mathbf{A}}^T$ where $[T]_\beta = \mathbf{A}$.

Proof. Let $\beta = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ be any orthonormal basis for $\mathcal{V}$.

We know that $\langle \vec{v}_i, \vec{v}_j \rangle = \delta_{ij}$ for all i, j. Let us look at $\langle T(\vec{v}_i), \vec{v}_j \rangle = \langle \vec{v}_i T^*(\vec{v}_j) \rangle$ from the definition of the adjoint linear operator.

Suppose we define $\mathbf{A} = [T]_\beta$ be a matrix with elements a_{ki}, where column i is $T(\vec{v}_i)$. Similarly, suppose we define $\mathbf{B} = [T^*]_\beta$ be a matrix with elements b_{kj}, where column j is $T^*(\vec{v}_j)$.

If we now write out the images, we have

$$\langle T(\vec{v}_i), \vec{v}_j \rangle = \left\langle \sum_k a_{ki} \vec{v}_k, \vec{v}_j \right\rangle = a_{ji} \langle \vec{v}_j, \vec{v}_j \rangle = a_{ji} \qquad \langle \vec{v}_i, T^*(\vec{v}_j) \rangle = \left\langle \vec{v}_i, \sum_k b_{kj} \vec{v}_k \right\rangle = \overline{b}_{ij} \langle \vec{v}_i, \vec{v}_i \rangle = b_{ij}$$

Here, the last equalities come from the fact that the only nonzero inner products are when $k = j$ and $k = i$ respectively (we conjugate b_{ij} because it comes from the second operand). Further, since these basis vectors are orthonormal, the last inner products evaluate to 1.

This means that $a_{ji} = \overline{b}_{ij}$ and $b_{ij} - \overline{a}_{ji}$; in other words, $\mathbf{B} = \overline{\mathbf{A}}^T$ and $[T^*]_\beta = \overline{\mathbf{A}}^T$. □

As such, we can conclude that T^* is unique (if it exists) and linear.

Lemma 34.7

Conversely, if we define T^* by $[T^*]_\beta = \overline{\mathbf{A}}^T$ for some orthonormal basis $\beta = \{\vec{v}_1, \dots, \vec{v}_n\}$, then $\langle T(\vec{x}), \vec{y} \rangle = \langle \vec{x}, T^*(\vec{y}) \rangle$ for all $\vec{x}, \vec{y} \in \mathcal{V}$.

Proof. We know that $\langle T(\vec{v}_i), \vec{v}_j \rangle = \langle \vec{v}_i, T^*(\vec{v}_j) \rangle$ for basis vectors $\vec{v}_i$ and $\vec{v}_j$; we can just backtrack from the previous lemma.

To extend this for all $\vec{x}, \vec{y} \in \mathcal{V}$, we can look at $\vec{x} = \sum_i c_i \vec{v}_i$ and $\vec{y} = \sum_j d_j \vec{v}_j$. We have

$$\langle T(\vec{x}), \vec{y} \rangle = \left\langle T\left(\sum_i c_i \vec{v}_i\right), \sum_j d_j \vec{v}_j \right\rangle$$

$$\begin{aligned}
&= \left\langle \sum_i c_i T(\vec{\boldsymbol{v}}_i), \sum d_j \vec{\boldsymbol{v}}_j \right\rangle \\
&= \sum_{i,j} c_i \bar{d}_j \langle T(\vec{\boldsymbol{v}}_i), \vec{\boldsymbol{v}}_j \rangle \\
&= \sum_{i,j} c_i \bar{d}_j \langle \vec{\boldsymbol{v}}_j, T^*(\vec{\boldsymbol{v}}_j) \rangle && \text{(works for basis)} \\
&= \left\langle \sum_i c_i \vec{\boldsymbol{v}}_i, T^*\left(\sum_i d_j \vec{\boldsymbol{v}}_j \right) \right\rangle \\
&= \langle \vec{\boldsymbol{x}}, T^*(\vec{\boldsymbol{y}}) \rangle
\end{aligned}$$

□

Our grand conclusion is that if $\mathcal{V}$ is finite dimensional, for any linear operator $T: \mathcal{V} \to \mathcal{V}$, there exists a unique $T^*: \mathcal{V} \to \mathcal{V}$ with matrix adjoint to T's in any orthonormal basis β; $[T^*]_\beta = [T]_\beta^*$.

An ultimate abstraction here is to look at the transformation $*: \mathcal{L}(\mathcal{V}, \mathcal{V}) \to \mathcal{L}(\mathcal{V}, \mathcal{V})$. That is, $*$ maps $T \mapsto T^*$. Is $*$ linear itself?

It turns out that it is almost linear; it's "conjugate-linear". For linear operators $T_1, T_2: \mathcal{V} \to \mathcal{V}$ and $\alpha, \beta \in \mathbb{C}$, we have

$$\left(\alpha T_1 + \beta T_2\right)^* = \bar{\alpha} T_1^* + \bar{\beta} T_2^*.$$

We also have $(T_1 \circ T_2)^* = T_2^* \circ T_1^*$; on the level of matrices, we have

$$(\mathbf{AB})^* = \overline{(\mathbf{AB})^T} = \overline{\mathbf{B}^T \mathbf{A}^T} = \overline{\mathbf{B}^T} \cdot \overline{\mathbf{A}^T} = \mathbf{B}^* \mathbf{A}^*.$$

Some good news is that $T^{**} = T$ and $\mathcal{I}^* = \mathcal{I}$ ($*$ preserves the unit).

11/19/2021

Lecture 35

Least Squares

The question we want to answer today is given a set of data points, we want to find the "best fit" linear function $f(x) = ax + b$ to those points.

Here is our setup.

We define the *residuals* to be the vertical differences from our projected $f(x_i)$'s to the true y_i's. Our goal will be to minimize the square sum of the residuals:

$$S = \sqrt{\sum_{i=1}^{n} (f(x_i) - y_i)^2}.$$

One question comes to mind; why not minimize the sum of the residuals $L_1 = \sum(f(x_i) - y_i)$? The problem here is that this value could evaluate to zero, yet there are points far from the line.

A followup would be to minimize the sum of absolute residuals $L_2 = \sum\left|f(x_i) - y_i\right|$? The problem here is that this expression is very hard to work with. Further, L_2 does not translate to distance in higher dimensions. Our original S is indeed the distance in higher dimensions.

Here are the different variables we'll be working with.

- S is the distance from a point to a subspace; in other words, we'll use the generalized Pythagorean theorem
- $\vec{\boldsymbol{x}}$ is our input data vector of x-coordinates

- $\vec{\boldsymbol{y}}$ is our output data vector of $\vec{\boldsymbol{y}}$-coordinates
- $\vec{\boldsymbol{w}} = \vec{\boldsymbol{f}}(x_i)$ is our vector of projected y_i's

In a general form, our $\vec{\boldsymbol{w}}$ can be written as

$$\vec{\boldsymbol{w}} = \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{bmatrix} = a \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + b \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

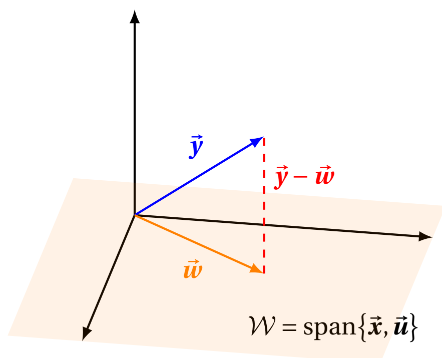
That is, $\vec{\boldsymbol{w}}$ is some linear combination of $\vec{\boldsymbol{x}}$ and $\vec{\boldsymbol{u}}$, where $\vec{\boldsymbol{u}}$ is a vector of 1's, $\vec{\boldsymbol{w}} = a\vec{\boldsymbol{x}} + b\vec{\boldsymbol{u}}$.

We can also rewrite S as

$$S = \sqrt{\sum_{i=1}^{n} (f(x_i) - y_i)^2} = \left\| \vec{\boldsymbol{w}} - \vec{\boldsymbol{y}} \right\| = \left\| a\vec{\boldsymbol{x}} + b\vec{\boldsymbol{u}} - \vec{\boldsymbol{y}} \right\|.$$

This is what we're trying to minimize. Note that all of the vectors $\vec{\boldsymbol{x}}$, $\vec{\boldsymbol{u}}$, and $\vec{\boldsymbol{y}}$ are fixed; a and b are varying, and our goal is to find these values.

Geometrically, we have



Note that $\vec{\boldsymbol{y}}$ may not be in $\mathcal{W}$, which means that solving the equality $a\vec{\boldsymbol{x}} + b\vec{\boldsymbol{u}} = \vec{\boldsymbol{y}}$ may not be possible.

In general, $\vec{\boldsymbol{y}} - \vec{\boldsymbol{w}} \perp \mathcal{W}$, and $\vec{\boldsymbol{w}} = \operatorname{proj}_{\mathcal{W}}(\vec{\boldsymbol{y}}) = a\vec{\boldsymbol{x}} + b\vec{\boldsymbol{u}}$.

There is another problem here; $\{\vec{\boldsymbol{x}}, \vec{\boldsymbol{u}}\}$ may not be orthonormal bases, meaning we cannot use our shortcut formulas for projections. This means that we need to derive formulas for projections without orthonormal bases.

35.1 Least Squares Formula

We now solve for a, b, and in the process S as well.

We introduce a matrix $\mathbf{A}$ as

$$\mathbf{A} = \begin{bmatrix} \vec{\boldsymbol{x}} & \vec{\boldsymbol{u}} \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}.$$

This means that we have

$$\vec{\boldsymbol{w}} - \mathbf{A} \begin{bmatrix} a \\ b \end{bmatrix}.$$

We want $\vec{\boldsymbol{y}} - \vec{\boldsymbol{w}} \perp \mathcal{W}$; to be perpendicular to $\mathcal{W}$, we need to be perpendicular to a spam. That is, we must have $\vec{\boldsymbol{y}} - \vec{\boldsymbol{w}} \perp \vec{\boldsymbol{x}}, \vec{\boldsymbol{u}}$.

Specifically, we can rewrite this as $\left\langle \vec{\boldsymbol{x}}, \vec{\boldsymbol{y}} - \vec{\boldsymbol{w}} \right\rangle = 0$ and $\left\langle \vec{\boldsymbol{u}}, \vec{\boldsymbol{y}} - \vec{\boldsymbol{w}} \right\rangle = 0$. This gives us

$$\left\langle \vec{\boldsymbol{x}}, \vec{\boldsymbol{y}} \right\rangle = \langle \vec{\boldsymbol{x}}, \vec{\boldsymbol{w}} \rangle$$
$$\left\langle \vec{\boldsymbol{u}}, \vec{\boldsymbol{y}} \right\rangle = \langle \vec{\boldsymbol{u}}, \vec{\boldsymbol{w}} \rangle$$

If we combine these two equations, we can see that the above is equivalent to

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \overline{y}_1 \\ \overline{y}_2 \\ \vdots \\ \overline{y}_n \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \overline{w}_1 \\ \overline{w}_2 \\ \vdots \\ \overline{w}_n \end{bmatrix}.$$

Taking out the conjugates, and moving them to the left matrices, we end up with

$$\begin{bmatrix} \overline{x}_1 & \overline{x}_2 & \cdots & \overline{x}_n \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \overline{x}_1 & \overline{x}_2 & \cdots & \overline{x}_n \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}.$$

This is exactly the equation $\mathbf{A}^* \vec{\boldsymbol{w}} = \mathbf{A}^* \vec{\boldsymbol{y}}$. However, we can rewrite $\vec{\boldsymbol{w}}$ again to give

$$\boxed{\mathbf{A}^*\mathbf{A} \begin{bmatrix} a \\ b \end{bmatrix} = \mathbf{A}^* \vec{\boldsymbol{y}}}.$$

This is what we call the *normal equation*.

If $\vec{\boldsymbol{y}} \in \mathcal{W} = \operatorname{Im} \mathbf{A}$, then all we need to do is solve $\mathbf{A} \begin{bmatrix} a \\ b \end{bmatrix} = \vec{\boldsymbol{y}}$. Doing it with the normal equation should give us the same solution.

However, $\mathbf{A}^*$ makes it possible to always solve the normal equation for our best fit line.

35.2 Applications

35.2.1 Statistics

We denote $\overline{x}$ as our arithmetic mean $\frac{1}{n}\sum x_i$, and $\overline{y}$ as our arithmetic mean $\frac{1}{n}\sum y_i$. Further, we let $\overline{x^2} = \frac{1}{n}\sum x_i^2$ and $\overline{xy} = \frac{1}{n}\sum x_i y_i$. This appears in our normal equation

$$\underbrace{\begin{bmatrix} \overline{x^2} & \overline{x} \\ \overline{x} & 1 \end{bmatrix}}_{\mathbf{A}^*\mathbf{A}} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \overline{xy} \\ \overline{y} \end{bmatrix}.$$

If we solve this, we have our formulas

$$a = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2} \qquad\qquad b = \overline{y} - a\overline{x}$$

35.2.2 Multivariable Calculus

If we look at S^2, we can see that it is a quadratic function of a and b; that is, $S^2 = F(a,b) = \sum_{i=1}^n (ax_i + b - y_i)^2$.

To solve the optimal result, we would like to solve for $\frac{\partial F}{\partial a} = 0$ and $\frac{\partial F}{\partial b} = 0$. This is a linear system; the squares get brought down, and we should get the same solution.

35.2.3 Probability

The correlation coefficient $r = \cos\alpha = \frac{\langle\cdot,\cdot\rangle}{\|\cdot\|\|\cdot\|}$ describes how correlated the two data streams are.

It turns out that the line we found is the most likely line; in our data points X and Y, the data points will most likely fall on the line $Y = aX + b + N$, where N represents our noise, assuming this noise is approximately normal with mean 0.

The outcome of all of this is that we can use the method of maximal likelihood to get values for a and b.

35.3 Generalizations

What if we would like to find the best fit parabola: $y = ax^2 + bx + c$.

Our $\mathbf{A}$ now becomes

$$\mathbf{A} = \begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ \vdots & \vdots & \vdots \\ x_n^2 & x_n & 1 \end{bmatrix}.$$

Everything else is still exactly the same; our matrix just becomes slightly larger, and we still solve $\mathbf{A}^*\mathbf{A}\vec{\boldsymbol{a}} = \mathbf{A}^*\vec{\boldsymbol{y}}$.

Let we look deeper at $\mathbf{A}^*\mathbf{A}$.

Lemma 35.1

We can see that $\mathbf{A}^*\mathbf{A}$ is a square matrix (i.e. multiplying $n \times m$ with $m \times n$ gives us $n \times n$), and specifically a hermitian matrix over $\mathbb{C}$ or a symmetric matrix over $\mathbb{R}$.

Proof. We have

$$(\mathbf{A}^*\mathbf{A})^* = (\mathbf{A}^*)(\mathbf{A}^{**}) = \mathbf{A}^*\mathbf{A}.$$

□

Definition 35.2: Self-Adjoint/Hermitian Matrices

We define a matrix $\mathbf{A}_{n\times n}$ to be *self-adjoint* (hermitian) if $\mathbf{A}^* = \mathbf{A}$.

Over $\mathbb{R}$, self-adjoint matrices are just symmetric matrices (that is, $\mathbf{A}^T = \mathbf{A}$).

We can see that

- Symmetric over $\mathbb{R} \implies$ hermitian over $\mathbb{C}$
- Hermitian over $\mathbb{C} \not\Longrightarrow$ symmetric over $\mathbb{R}$
- Symmetric over $\mathbb{C} \not\Longrightarrow$ hermitian over $\mathbb{C}$
- Hermitian over $\mathbb{C} \not\Longrightarrow$ symmetric over $\mathbb{C}$.

Note that self-adjoint complex matrices must have real entries along the diagonal (as their conjugate must be equal to themselves).

11/22/2021

Lecture 36

Symmetric and Hermitian Operators

Theorem 36.1: Diagonalizability of Self-Adjoint Matrices

Every self-adjoint matrix $\mathbf{A}$ is diagonalizable over $\mathbb{R}$.

That is, for any self-adjoint matrix $\mathbf{A}$, there exists a diagonal matrix $\mathbf{D}_{n\times n}$ over $\mathbb{R}$ and an invertible matrix $\mathbf{U}$ over $\mathbb{R}$ or $\mathbb{C}$ such that $\mathbf{A} = \mathbf{UDU}^{-1}$.

Moreover, $\mathbf{U}$ can be chosen to be orthogonal or unitary.

Definition 36.2: Orthogonal and Unitary Matrices

A matrix $\mathbf{U} \in \mathcal{M}_{n\times n}$ is orthogonal over $\mathbb{R}$ or unitary over $\mathbb{C}$ if its columns (or equivalently, rows) form an orthonormal basis for $\mathbb{R}^n$ (or $\mathbb{C}^n$).

Example 36.3

Here, we have

$$\mathbf{U} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \qquad \mathbf{U} = \begin{bmatrix} i & 0 \\ 0 & 1 \end{bmatrix}$$

both of which are orthogonal/unitary matrices.

Definition 36.4: Self-Adjoint Operator

Suppose we have an inner product space $\mathcal{V}_{\langle\cdot,\cdot\rangle}$. We define a linear operator $T : \mathcal{V} \to \mathcal{V}$ to be a *self-adjoint operator* if $T = T^*$, i.e.

$$\langle T(\vec{\boldsymbol{v}}), \vec{\boldsymbol{w}}\rangle = \langle \vec{\boldsymbol{v}}, T(\vec{\boldsymbol{w}})\rangle, \qquad \forall \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}} \in \mathcal{V}.$$

Example 36.5

For example, the identity operator $\mathcal{I}_{\mathcal{V}}$ and the zero operator $0_{\mathcal{V}}$ are both self-adjoint operators.

Recall that T^* is adjoint to T if and only if

$$\langle T(\vec{\boldsymbol{v}}), \vec{\boldsymbol{w}}\rangle = \left\langle \vec{\boldsymbol{v}}, T^*(\vec{\boldsymbol{w}})\right\rangle, \qquad \forall \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}} \in \mathcal{V}.$$

that is, $[T^*]_\beta = [T]_\beta^*$ in some (and hence all) orthonormal basis β.

Corollary 36.6

Over an inner product space $\mathcal{V}$, a linear operator $T : \mathcal{V} \to \mathcal{V}$ is self-adjoint if and only if $[T]_\beta^* = [T]_\beta$; that is, T has a self-adjoint matrix $\mathbf{A}$ in some (and hence all) orthonormal basis β.

Further, we say T is symmetric if $\mathbf{A}^T = \mathbf{A}$ over $\mathbb{R}$, and we say T is hermitian if $\mathbf{A}^* = \mathbf{A}$ over $\mathbb{C}$. (In some/any/all orthonormal bases)

Our next question: what can we say about the eigenvalues of a self-adjoint T?

Theorem 36.7: Real Eigenvalues of a Self-Adjoint Matrix

Let $\mathbf{A}$ be a symmetric/hermitian (i.e. self-adjoint) matrix. Then all eigenvalues of $\mathbf{A}$ are real and $\text{char}_{\mathbf{A}}(\lambda)$ splits over $\mathbb{R}$.

Proof. Consider a matrix $\mathbf{A} \in \mathcal{M}_{n\times n}(\mathbb{C})$. We know that $\text{char}_{\mathbf{A}}(\lambda)$ splits over $\mathbb{C}$ with $\lambda_i \in \mathbb{C}$.

Now, let our inner product be our standard inner product over $\mathbb{C}$.

Here, we know that $\mathbf{A}$ corresponds to a hermitian operator T. That is, $[T]_e = \mathbf{A}$ and $\mathbf{A}^* = \mathbf{A}$. What does this mean for the linear operator? We have that

$$\langle T(\vec{\boldsymbol{v}}), \vec{\boldsymbol{w}}\rangle = \langle \vec{\boldsymbol{v}}, T(\vec{\boldsymbol{w}})\rangle, \qquad \forall \vec{\boldsymbol{v}}, \vec{\boldsymbol{w}} \in \mathbb{C}^n.$$

Now, let us bring in the eigenvectors.

Suppose $\vec{\boldsymbol{v}}$ is an eigenvector of $\mathbf{A}$ corresponding to λ_i. We then have

$$\begin{aligned}
\langle \mathbf{A}(\vec{\boldsymbol{v}}), \vec{\boldsymbol{v}}\rangle &= \langle \vec{\boldsymbol{v}}, \mathbf{A}(\vec{\boldsymbol{v}})\rangle \\
\langle \lambda_i \vec{\boldsymbol{v}}, \vec{\boldsymbol{v}}\rangle &= \langle \vec{\boldsymbol{v}}, \lambda_i \vec{\boldsymbol{v}}\rangle \\
\lambda_i \langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{v}}\rangle &= \overline{\lambda}_i \langle \vec{\boldsymbol{v}}, \vec{\boldsymbol{v}}\rangle \\
\left(\lambda_i - \overline{\lambda}\right) \|\vec{\boldsymbol{v}}\|^2 &= 0
\end{aligned}$$

We know that eigenvectors cannot be zero, so $\|\vec{\boldsymbol{v}}\|^2 \neq 0$. As such, the first term must be zero; we must have $\lambda_i = \overline{\lambda_i}$.

Hence, λ_i must be real, and $\text{char}_{\mathbf{A}}(\lambda)$ must split over $\mathbb{R}$. □

We've established that we have all the eigenvalues, but we may not have enough eigenvectors to diagonalize the matrix.

Theorem 36.8: Orthogonal Eigenspaces for Hermitian Matrices

Suppose $\mathbf{A}$ is a symmetric/hermitian matrix. If $\vec{\boldsymbol{v}}_1$ and $\vec{\boldsymbol{v}}_2$ are eigenvectors correspond to $\lambda_1 \neq \lambda_2$ ($\in \mathbb{R}$), then $\vec{\boldsymbol{v}}_1 \perp \vec{\boldsymbol{v}}_2$ (with respect to the standard inner product).

Proof. We have

$$\begin{aligned}
\langle \mathbf{A}\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2\rangle &= \langle \vec{\boldsymbol{v}}_1, \mathbf{A}\vec{\boldsymbol{v}}_2\rangle \\
\langle \lambda_1 \vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2\rangle &= \langle \vec{\boldsymbol{v}}_1, \lambda_2 \vec{\boldsymbol{v}}_2\rangle \\
\lambda_1 \langle \vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2\rangle &= \overline{\lambda}_2 \langle \vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2\rangle \\
\left(\lambda_1 - \overline{\lambda_2}\right) \langle \vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2\rangle &= 0
\end{aligned}$$

We know that $\lambda_1, \lambda_2 \in \mathbb{R}$ and that $\lambda_1 \neq \lambda_2$, so the first term cannot be zero. As such, $\langle \vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2\rangle = 0$ and we know that $\vec{\boldsymbol{v}}_1 \perp \vec{\boldsymbol{v}}_2$.

This further leads to another consequence; the entire eigenspaces must be perpendicular for distinct eigenvalues; $E_{\lambda_1} \perp E_{\lambda_2}$ for $\lambda_1 \neq \lambda_2$. □

Let us now translate this to linear operators.

Corollary 36.9

If we have a symmetric/hermitian operator T in any inner product space $\mathcal{V}$, then all eigenvalues of T are real, $\text{char}_T(\lambda)$ splits over $\mathbb{R}$, any eigenvectors $\vec{\boldsymbol{v}}_1$ and $\vec{\boldsymbol{v}}2$ corresponding to $\lambda_1 \neq \lambda_2$ are orthogonal.

Recall that the orthogonal complement of a subspace $\mathcal{W} \subsetneq \mathcal{V}$ of an inner product space $\mathcal{V}$ is defined to be

$$\mathcal{W}^\perp = \left\{\vec{\boldsymbol{v}} \in \mathcal{V} \mid \vec{\boldsymbol{v}} \perp \mathcal{W}\right\}.$$

We also have that $\mathcal{W}^\perp$ is a subspace.

Further, we can extend any orthonormal basis of $\mathcal{W}$ to an orthonormal basis for $\mathcal{V}$, and the vectors we use to extend the basis forms a basis for $\mathcal{W}^\perp$. That is, if $\dim \mathcal{V} < \infty$, then $\mathcal{W} \oplus \mathcal{W}^\perp = \mathcal{V}$, and $\dim \mathcal{W} + \dim \mathcal{W}^\perp = \dim \mathcal{V}$.

Corollary 36.10

If we have a symmetric/hermitian operator T on an inner product space $\mathcal{V}$, where $\mathcal{Q}$ is a T-invariant subspace of $\mathcal{V}$, we have that $\mathcal{W}^\perp$ is also a T-invariant subspace of $\mathcal{V}$.

Proof. Let $\vec{\boldsymbol{v}} \in \mathcal{W}^\perp$. We want to show that $T(\vec{\boldsymbol{v}}) \in \mathcal{W}^\perp$. Let $\vec{\boldsymbol{w}} \in \mathcal{W}$; it is equivalent to show that $T(\vec{\boldsymbol{v}}) \perp \vec{\boldsymbol{w}}$.

We have

$$\langle T(\vec{\boldsymbol{v}}), \vec{\boldsymbol{w}}\rangle = \langle \vec{\boldsymbol{v}}, T(\vec{\boldsymbol{w}})\rangle.$$

Notice that $T(\vec{\boldsymbol{w}}) \in \mathcal{W}$, while $\vec{\boldsymbol{v}} \in \mathcal{W}^\perp$, and as such $\langle \vec{\boldsymbol{v}}, T(\vec{\boldsymbol{w}})\rangle = 0$. This means that $\mathcal{W}^\perp$ is T-invariant, as $T(\vec{\boldsymbol{v}}) \perp \vec{\boldsymbol{w}}$. □

As such, because $\mathcal{V}$ can be written as a direct sum of two orthonormal and T-invariant subspaces, we can see that

$$[T]_{\beta_1 \cup \beta_2} = \begin{bmatrix} [T_{\mathcal{W}}]_{\beta_1} & 0 \\ 0 & [T_{\mathcal{W}^\perp}]_{\beta_2} \end{bmatrix}.$$

Now, we will prove the earlier theorem.

Proof. Suppose $\dim \mathcal{V} = n$, and we will proceed by induction on n.

In the base case of $n = 1$, there is nothing to show. Let us assume taht the claim is true for all $\leq n-1$.

We want to show the case for n, i.e. $\dim \mathcal{V} = n$.

Starting with one eigenvalue $\lambda_1 \in \mathbb{R}$, let $\vec{\boldsymbol{w}}1$ be an eigenvector of T corresponding to λ_1, and let $\|\vec{\boldsymbol{w}}_1\| = 1$.

Let $\mathcal{W} = \text{span}\left\{\vec{\boldsymbol{w}}_1\right\}$. This subspace is a 1-dimensional T-invariant subspace of $\mathcal{V}$.

Taking its orthogonal complement $\mathcal{W}^\perp$, we know that this must be a $n-1$ dimensional T-invariant subspace of $\mathcal{V}$.

If we restrict T to $\mathcal{W}^\perp$, we know that $T_{\mathcal{W}^\perp}$ must also be self-adjoint (as this only depends on the inner product, not the size of the vector space).

By the IH, we then know that we can diagonalize $T_{\mathcal{W}^\perp}$ with an orthonormal basis $\beta_2 = \{\vec{\boldsymbol{w}}_2, \ldots, \vec{\boldsymbol{w}}_n\}$ for $\mathcal{W}^\perp$ such that $[T_{\mathcal{W}^\perp}]_{\beta_2}$ is diagonal.

Therefore, if we add $\vec{\boldsymbol{w}}_1$ to this basis to form $\beta = \{\vec{\boldsymbol{w}}_1, \vec{\boldsymbol{w}}_2, \ldots, \vec{\boldsymbol{w}}_n\}$, we have an orthonormal basis for $\mathcal{V} = \mathcal{W} \oplus \mathcal{W}^\perp$, in which $[T]_\beta$ is diagonal:

$$[T]_\beta = \begin{bmatrix} [T_{\mathcal{W}}]_{\beta_1} & 0 \\ 0 & [T_{\mathcal{W}^\perp}]_{\beta_2} \end{bmatrix}.$$

□

Theorem 36.11: Spectral Theorem

Let $\mathcal{V}$ be an inner product space, with $T: \mathcal{V} \to \mathcal{V}$ a self-adjoint operator.

Suppose

- $\lambda_1, \dots, \lambda_k$ are distinct eigenvalues over $\mathbb{R}$
- E_{λ_i} is an eigenspace corresponding to λ_i
- T_i is the orthogonal projection $\mathcal{V} \to E_{\lambda_i}$

We then have $\mathcal{V} = E_{\lambda_1} \oplus E_{\lambda_2} \oplus \cdots \oplus E_{\lambda_k}$. Further, $E_{\lambda_i}^{\perp}$ is the direct sum of the other eigenspaces.

Further, $T_i T_j = \delta_{ij} T_i = \begin{cases} 0 & \text{if } i \neq j \\ T_i & \text{if } i = j \end{cases}$ We also have $\mathcal{I} = T_1 + T_2 + \cdots + T_k$, and $T = \lambda_1 T_1 + \lambda_2 + T_2 + \cdots + \lambda_k T_k$.

11/29/2021

Lecture 37

Orthogonal/Unitary Operators

Recall that the heart of the spectral theorem says that T being self-adjoint is sufficient to conclude that T is diagonalizable in an orthonormal basis.

Our question is whether this is a necessary condition. Over $\mathbb{R}$, it turns out that it is necessary and sufficient; if T is diagonalizable in some orthonormal basis, then T is self-adjoint. However, it is not necessary over $\mathbb{C}$; it turns out that it is sufficient and necessary to be a *normal* operator over $\mathbb{C}$.

Let us go through the motivation of this.

Recall that a linear operator $T: \mathcal{V} \to \mathcal{V}$ preserves the vector space structure—they preserve linear combinations.

A followup question would be: what linear operators preserve the inner product structure? This means that we must have $\left\langle T(\vec{\boldsymbol{x}}), T(\vec{\boldsymbol{y}})\right\rangle = \left\langle \vec{\boldsymbol{x}}, \vec{\boldsymbol{y}}\right\rangle$ for all $\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}} \in \mathcal{V}$. Suppose $\vec{\boldsymbol{x}} = \vec{\boldsymbol{y}}$; that means that $\langle T(\vec{\boldsymbol{x}}), T(\vec{\boldsymbol{x}})\rangle = \langle \vec{\boldsymbol{x}}, \vec{\boldsymbol{x}}\rangle$, which implies that $\|T(\vec{\boldsymbol{x}})\| = \|\vec{\boldsymbol{x}}\|$; T preserves the length of vectors.

Let us consider such an operator that preserves length; we must also preserve angles as a result! Recall that $\cos\alpha = \frac{\langle \vec{\boldsymbol{x}}, \vec{\boldsymbol{y}}\rangle}{\|\vec{\boldsymbol{x}}\| \cdot \|\vec{\boldsymbol{y}}\|} \in [-1, 1]$. Since T preserves lengths and inner products, we then must have $\cos\alpha = \frac{\langle T(\vec{\boldsymbol{x}}), T(\vec{\boldsymbol{y}})\rangle}{\|T(\vec{\boldsymbol{x}})\| \cdot \|T(\vec{\boldsymbol{y}})\|}$; as such, T must also preserve angles.

What about the converse? Which of these properties is enough for the rest? We can clearly see that preserving inner products implies the preservation of lengths and angles, since that's how we derived everything.

It turns out that preserving the length is also enough to ensure preservation of inner products and angles.

Lemma 37.1

If $T: \mathcal{V} \to \mathcal{V}$ is a linear operator, and T preserves lengths, then T preserves inner products (and hence angles as well).

Proof. Suppose we have $\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}} \in \mathcal{V}$. Since T preserves lengths, we know

$$\left\langle T(\vec{\boldsymbol{x}} + \vec{\boldsymbol{y}}), T(\vec{\boldsymbol{x}} + \vec{\boldsymbol{y}})\right\rangle = \left\langle \vec{\boldsymbol{x}} + \vec{\boldsymbol{y}}, \vec{\boldsymbol{x}} + \vec{\boldsymbol{y}}\right\rangle$$
$$\langle T(\vec{\boldsymbol{x}}), T(\vec{\boldsymbol{x}})\rangle + \left\langle T(\vec{\boldsymbol{y}}), T(\vec{\boldsymbol{y}})\right\rangle + \left\langle T(\vec{\boldsymbol{x}}), T(\vec{\boldsymbol{y}})\right\rangle + \left\langle T(\vec{\boldsymbol{y}}), T(\vec{\boldsymbol{x}})\right\rangle = \langle \vec{\boldsymbol{x}}, \vec{\boldsymbol{x}}\rangle + \left\langle \vec{\boldsymbol{y}}, \vec{\boldsymbol{y}}\right\rangle + \left\langle \vec{\boldsymbol{x}}, \vec{\boldsymbol{y}}\right\rangle + \left\langle \vec{\boldsymbol{y}}, \vec{\boldsymbol{x}}\right\rangle$$

Since T preserves lengths, we can cancel out the first two terms on both sides:

$$\langle T(\vec{x}), T(\vec{y})\rangle + \langle T(\vec{y}), T(\vec{x})\rangle = \langle \vec{x}, \vec{y}\rangle + \langle \vec{y}, \vec{x}\rangle$$

Over $\mathbb{R}$, we're actually done, since this means that

$$\langle T(\vec{x}), T(\vec{y})\rangle = \langle \vec{x}, \vec{y}\rangle$$

However, we need to be a little bit careful over $\mathbb{C}$; we'd end up with the conclusion that

$$\mathrm{Re}\langle \vec{x}, \vec{y}\rangle = \mathrm{Re}\langle T(\vec{x}), T(\vec{y})\rangle,$$

meaning the real parts of the inner products are preserved.

To make conclusions about the imaginary part, we'd instead start with $\vec{x} + i\vec{y}$ instead of just $\vec{x} + \vec{y}$; we'd end up with $\mathrm{Im}\langle \vec{x}, \vec{y}\rangle = \mathrm{Im}\langle T(\vec{x}), T(\vec{y})\rangle$ after a very similar computation.

As such, since T preserves both the real and imaginary parts of the inner product, it must preserve the inner product.

Since T preserves the inner product, then we must also preserve angles, as derived before. $\square$

We just justified the definition of orthogonal/unitary linear operators:

Definition 37.2: Orthogonal/Unitary Linear Operator

Let $T: \mathcal{V} \to \mathcal{V}$ be a linear operator on an inner product space with finite dimension over $\mathbb{R}$ or $\mathbb{C}$.

If T preserves length—that is, $\|T(\vec{x})\| = \|\vec{x}\|$ for all $\vec{x}$—then T is orthogonal (over $\mathbb{R}$) or unitary (over $\mathbb{C}$).

One caveat is that preservation of angles *does not* necessarily imply the preservation of lengths and inner products—take $T = 2\mathcal{I}$; scaling does not preserve lengths, but do preserve angles.

Our next question: what are the matrices of unitary/orthogonal operators?

Let $\beta = \{\vec{v}_i\}_1^n$ be an orthonormal basis for $\mathcal{V}$. We then have $T(\beta) = \{T(\vec{v}_i)\}_1^n$ is another orthonormal basis for $\mathcal{V}$; this is easily derived by the fact that $\langle T(\vec{v}_i), T(\vec{v}_j)\rangle = \langle \vec{v}_i, \vec{v}_j\rangle = \delta_{ij}$.

This means that the columns of the matrix $[T]_\beta$ are also orthonormal; they're just the images $[T(\vec{v}_i)]_\beta$. As such, we also have orthonormal rows.

Further, this implies that $[T]_\beta^*[T]_\beta = [T]_\beta[T]_\beta^* = \mathbf{I}_n$; this leads us to the definition of unitary/orthogonal matrices:

Definition 37.3: Othogonal/Unitary Matrices

A matrix $\mathbf{A}_{n\times n}$ over $\mathbb{C}$ or $\mathbb{R}$ is unitary/orthogonal if $\mathbf{A}^{-1} = \mathbf{A}^*$.

Theorem 37.4: Orthogonal in Orthonormal Bases

A linear operator $T: \mathcal{V} \to \mathcal{V}$ over an inner product space is unitary/orthogonal if and only if $[T]_\beta = \mathbf{A}$ is unitary/orthogonal in any orthonormal basis (and thus all orthonormal bases).

Proof. ($\Longrightarrow$) We just proved the forward direction previously.

($\Longleftarrow$) If $\mathbf{A}$ is unitary/orthogonal, then let us consider $\mathbf{A} = [T]_\beta$, the linear operator over some orthonormal basis β. This means that

$$\langle T(\vec{\boldsymbol{v}}_i), T(\vec{\boldsymbol{v}}_j)\rangle = \delta_{ij} = \langle \vec{\boldsymbol{v}}_i, \vec{\boldsymbol{v}}_j\rangle,$$

because the matrix $\mathbf{A}$ over an orthonormal basis has orthonormal columns.

We can further extend this to show that $\langle T(\vec{\boldsymbol{x}}), T(\vec{\boldsymbol{y}})\rangle = \langle \vec{\boldsymbol{x}}, \vec{\boldsymbol{y}}\rangle$ if we extend linearly; we'd just expand $\vec{\boldsymbol{x}}$ and $\vec{\boldsymbol{y}}$ as a linear combination of basis vectors $\vec{\boldsymbol{v}}_i$'s.

As such, T is an orthogonal/unitary operator. □

What is the big picture?

We have three kinds of matrices (over $\mathbb{C}$):

- Unitary/orthogonal matrices ($\mathbf{A}\mathbf{A}^* = \mathbf{I}_n$; $\mathbf{A}^* = \mathbf{A}^{-1}$)
- Normal matrices ($\mathbf{A}\mathbf{A}_6 * = \mathbf{A}^*\mathbf{A}$)
- Self-adjoint matrices ($\mathbf{A}^* = \mathbf{A}$)

What are the implications?

- Unitary/orthogonal $\implies$ normal
- Self adjoint $\implies$ normal

Further, we know that all of these three kinds of matrices must be diagonalizable in some orthonormal basis over $\mathbb{C}$ by the spectral theorem—it says that all normal matrices are diagonalizable.

Over $\mathbb{R}$, we have something slightly different; we only have symmetric matrices over $\mathbb{R}$ (i.e. $\mathbf{A}^T = \mathbf{A}$). We can see that symmetric matrices must also be diagonalizable over $\mathbb{R}$, by the same spectral theorem.

Let us look at normal matrices over $\mathbb{C}$; we know that some matrices are unitary and some matrices are hermitian. What happens if T is both unitary and hermitian?

This means that $[T]_\beta$ is diagonalizable over some orthonormal basis, and must also be unitary over all orthonormal bases.

Since T is diagonalizable with real eigenvalues, and $\lambda_i^2 = 1$ (because T is unitary), we must have $\lambda_i = \pm 1$. That is, all eigenvalues are units.

As such, the diagonal entries of $[T]_\beta$ must all be ± 1, and zeroes elsewhere.

Definition 37.5: Rigid Motion

A function over a inner product space $f : \mathcal{V} \to \mathcal{V}$ is a *rigid motion* if it preserves distances:

$$\left\| f(\vec{\boldsymbol{x}}) - f(\vec{\boldsymbol{y}})\right\| = \left\|\vec{\boldsymbol{x}} - \vec{\boldsymbol{y}}\right\| \quad \forall \vec{\boldsymbol{x}}, \vec{\boldsymbol{y}} \in \mathcal{V}.$$

The reason why we do not specify linear is because translations are not linear, yet preserve distances.

Theorem 37.6: Rigid Motion as a Composition of Transformations

If $f : \mathcal{V} \to \mathcal{V}$ is a rigid motion, then there exists a unique orthogonal $T : \mathcal{V} \to \mathcal{V}$ and a unique translation $g : \mathcal{V} \to \mathcal{V}$ such that $T \circ g = f$.

12/1/2021

Lecture 38

Quadratic Forms and Positive Definite Quadratic Forms

Example 38.1

As motivation, consider the function $q(x_1, x_2) = 8x_1^2 - 4x_1x_2 + 5x_2^2$.

How would we find the global minimum/maximum, or show if it does not exist?

As a start, is $q(0,0)$ a global extremum? The multivariable calculus approach would be to take the partial derivatives:

$$\frac{\partial q}{\partial x_1} = 16x_1 - 4x_2 = 0$$
$$\frac{\partial q}{\partial x_2} = -4x_1 + 10x_2 = 0$$

We can see that $\frac{\partial q}{\partial x_1}(0,0) = 0 = \frac{\partial q}{\partial x_2}(0,0)$, and it turns out that this is the only common solution to the system.

In other words, we'd have $\nabla q(0,0) = 0$, which means that it is a potential extremum.

However, we'd also need to do something like the second derivative test, and do a bunch of other math to figure out whether it is actually a global min/max.

We'll be looking at a linear algebra approach to this problem. (Here, we will use the specific example to guide us.)

We have the function $q(x_1, x_2)$, and we'll introduce the vector $\vec{\boldsymbol{x}} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$.

We can then rewrite q as

$$q(x_1, x_2) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 8x_1 & -2x_2 \\ -2x_1 & 5x_2 \end{bmatrix}.$$

(Note that there are infinitely many possible terms on the anti-diagonal, but here we have split the $-4x_1x_2$ evenly.)

We can further rewrite this as

$$q(x_1, x_2) = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}^2 \begin{bmatrix} 8 & -2 \\ -2 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{\boldsymbol{x}}^T \mathbf{A} \vec{\boldsymbol{x}}.$$

Here, we have $\mathbf{A}$ as a symmetric square matrix. Notice that the choice of splitting $-4x_1x_2$ into two $-2x_1x_2$'s made this symmetric (it's the only choice we have).

Since $\mathbf{A}$ is symmetric, we can diagonalize it in some orthonormal basis β of $\mathbb{R}^2$.

We have the characteristic polynomial $\operatorname{char}_{\mathbf{A}}(\lambda) = \lambda^2 - 13\lambda + 36$, which means we have $\lambda_1 = 9$, $\lambda_2 = 4$.

The corresponding eigenvectors are $\vec{\boldsymbol{v}}_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ and $\vec{\boldsymbol{v}}_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, normalized to form our orthonormal eigenbasis β.

As such, we have $\mathbf{A} = \mathbf{Q}^{-1}\mathbf{D}\mathbf{Q} = \mathbf{Q}^T\mathbf{D}\mathbf{Q}$ where $\mathbf{D}$ is the diagonal matrix of eigenvalues, and $\mathbf{Q}$ is the orthonormal matrix of eigenvectors.

We know that for all $\vec{\boldsymbol{x}} \in \mathbb{R}^2$, we have $\vec{\boldsymbol{x}} = a_1 \vec{\boldsymbol{v}}1 + a_2 \vec{\boldsymbol{v}}_2$; plugging into $q(\vec{\boldsymbol{x}})$, we have

$$\begin{aligned}
\vec{\boldsymbol{x}}^T \mathbf{A} \vec{\boldsymbol{x}} &= (a_1 \vec{\boldsymbol{v}}_1 + a_2 \vec{\boldsymbol{v}}_2)^T (\mathbf{A}(a_1 \vec{\boldsymbol{v}}_1 + a_2 \vec{\boldsymbol{v}}_2)) \\
&= (a_1 \vec{\boldsymbol{v}}_1 + a_2 \vec{\boldsymbol{v}}_2)^T (a_1 \cdot 9 \vec{\boldsymbol{v}}_1 + a_2 \cdot 4 \vec{\boldsymbol{v}}_2) \\
&= 9a_1^2 + 4a_2^2 \geq 0
\end{aligned}$$

Here, lots of things cancel out because $\langle \vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2 \rangle = 0$ and $\langle \vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_1 \rangle = \langle \vec{\boldsymbol{v}}_2, \vec{\boldsymbol{v}}_2 \rangle = 1$.

Equivalently, we have $q(x_1, x_2) = [\vec{\boldsymbol{x}}]_\beta^T \begin{bmatrix} 9 & 0 \\ 0 & 4 \end{bmatrix} [\vec{\boldsymbol{x}}]_\beta$, because $\vec{\boldsymbol{x}}^T \mathbf{A} \vec{\boldsymbol{x}} = \vec{\boldsymbol{x}}^T \mathbf{Q}^T \mathbf{D} \mathbf{Q} \vec{\boldsymbol{x}} = (\mathbf{Q}\vec{\boldsymbol{x}})^T \mathbf{D} (\mathbf{Q}\vec{\boldsymbol{x}})$.

In conclusion, we know that $q(\vec{\boldsymbol{x}}) \geq 0$ for all $\vec{\boldsymbol{x}} \in \mathbb{R}^2$, with equality if and only if $a_1 = a_2 = 0$, which occurs if and only if $\vec{\boldsymbol{x}} = \vec{\mathbf{0}}$.

Hence, $q(0,0)$ is a global minimum of $q(x_1, x_2)$.

38.1 Generalizations

We can now generalize to $\mathbb{R}^n$.

Definition 38.2: Quadratic Form

A function $q(x_1, \dots, x_n) : \mathbb{R}^n \to \mathbb{R}$ is a *quadratic form* if it is a linear combination of $x_i x_j$ for $1 \leq i, j \leq n$):

$$q(x_1, \dots, x_n) = \sum_{i,j} \alpha_{ij} x_i x_j.$$

Theorem 38.3: Quadratic Form in Matrix Form

A quadratic form $q : \mathbb{R}^n \to \mathbb{R}$ can be written as

$$q(\vec{\boldsymbol{x}}) = \vec{\boldsymbol{x}}^T \mathbf{A} \vec{\boldsymbol{x}},$$

for a unique symmetric $n \times n$ matrix $\mathbf{A}$.

Proof. We've already shown that q can be written in this form, but we'll now prove that this matrix $\mathbf{A}$ is unique.

Suppose we plug in the ith standard basis vector $\vec{\boldsymbol{e}}_i$:

$$q(\vec{\boldsymbol{e}}_i) = \vec{\boldsymbol{e}}_i^T \mathbf{A} \vec{\boldsymbol{e}}_i = a_{ii}.$$

That is, $q(\vec{\boldsymbol{e}}_i)$ gives us the ith entry along the diagonal of $\mathbf{A}$—that is, we have the unique coefficient of x_i^2 (because $\vec{\boldsymbol{e}}_i$ represents x_i).

We could try the same thing with $\vec{\boldsymbol{e}}_i$ and $\vec{\boldsymbol{e}}_j$, but this would not work—we'd need to have the same $\vec{\boldsymbol{x}}$ on both sides of $\mathbf{A}$.

Hence, let us look at a linear combination of $\vec{\boldsymbol{e}}_i$ and $\vec{\boldsymbol{e}}_j$, i.e. $\vec{\boldsymbol{e}}_i + \vec{\boldsymbol{e}}_j$:

$$\begin{aligned} q(\vec{\boldsymbol{e}}_i + \vec{\boldsymbol{e}}_j) &= (\vec{\boldsymbol{e}}_i + \vec{\boldsymbol{e}}_j)^T \mathbf{A} (\vec{\boldsymbol{e}}_i + \vec{\boldsymbol{e}}_j) \\ &= \vec{\boldsymbol{e}}_i^T \mathbf{A} \vec{\boldsymbol{e}}_i + \vec{\boldsymbol{e}}_i^T \mathbf{A} \vec{\boldsymbol{e}}_j + \vec{\boldsymbol{e}}_j^T \mathbf{A} \vec{\boldsymbol{e}}_i + \vec{\boldsymbol{e}}_j^T \mathbf{A} \vec{\boldsymbol{e}}_j \\ &= q(\vec{\boldsymbol{e}}_i) + q(\vec{\boldsymbol{e}}_j) + a_{ij} + a_{ji} \end{aligned}$$

In order for $\mathbf{A}$ to be symmetric, we must have $a_{ij} = a_{ji} = \frac{1}{2}\big(q(\vec{\boldsymbol{e}}_i + \vec{\boldsymbol{e}}_j) - q(\vec{\boldsymbol{e}}_i) + q(\vec{\boldsymbol{e}}_j)\big)$.

As such, since both elements have a formula involving only q, if q is given, their values are fixed, and $\mathbf{A}$ is unique. $\square$

In practice, we won't be going through all of these caluclations to find $\mathbf{A}$; we can just split the terms into two parts, giving us

$$q(\vec{\boldsymbol{x}}) = \sum_{i=1}^n a_{ii} x_i^2 + \sum_{i \neq j} 2 a_{ij} x_i x_j.$$

Example 38.4

Let us take
$$q(x_1, x_2, x_3) = 9x_1^2 + 7x_2^2 + 3x_3^2 - 2x_1x_1 + 4x_1x_2 + 4x_1x_3 - 6x_2x_3.$$
We have $q(\vec{x}) = \vec{x}^T \mathbf{A}\vec{x}$, where
$$\mathbf{A} = \begin{bmatrix} 9 & -1 & 2 \\ -1 & 7 & -3 \\ 2 & -3 & 3 \end{bmatrix}.$$
We can now diagonalize $\mathbf{A}$ and solve.

Since $\mathbf{A}$ is symmetric over $\mathbb{R}$, we can write $\mathbf{A} = \mathbf{Q}^{-1}\mathbf{D}\mathbf{Q}$ in some orthonormal basis $\beta = \{\vec{v}_1, \dots, \vec{v}_n\}$ of $\mathbb{R}^n$, where $\mathbf{D}$ is our diagonal matrix of eigenvalues λ_i.

We can then write $q(\vec{x}) = [\vec{x}]_\beta^T \mathbf{D}[\vec{x}]_\beta$, and expanding it out again we have
$$q(\vec{x}) = \lambda_1 c_1^2 + \lambda_2 c_2^2 + \cdots + \lambda_n c_n^2,$$
where c_i is equal to the ith element of $[\vec{x}]_\beta$.

Notice that there are no mixed terms—we only have squares and their coefficients.

38.2 Efficient optimization

Our goal here is to find a way to efficiently determine the global minimum/maximum (or determine if it does not exist) for quadratic forms.

Definition 38.5: Definiteness

Suppose we have a quadratic form $q(\vec{x}) : \mathbb{R}^n \to \mathbb{R}$, or equivalently its corresponding matrix $\mathbf{A}$.

We define q (equivalently $\mathbf{A}$) to be

- *Positive definite* if $q(\vec{x}) > 0$ (equivalently $\vec{x}^T\mathbf{A}\vec{x} > 0$) for all $\vec{x} \neq \vec{\mathbf{0}}$.
- *Negative definite* if $q(\vec{x}) < 0$ (equivalently $\vec{x}^T\mathbf{A}\vec{x} < 0$) for all $\vec{x} \neq 0$.
- *Positive semi-definite* if $q(\vec{x}) \geq 0$ (equivalently $\vec{x}^T\mathbf{A}\vec{x} \geq 0$) for all $\vec{x} \neq 0$.
- *Negative semi-definite* if $q(\vec{x}) \leq 0$ (equivalently $\vec{x}^T\mathbf{A}\vec{x} \leq 0$) for all $\vec{x} \neq 0$.
- *Indefinite* if $q(\vec{x}) < 0$ and $q(\vec{y}) > 0$ (equivalently $\vec{x}^T\mathbf{A}\vec{x} < 0$ and $\vec{y}^T\mathbf{A}\vec{y} > 0$) for some $\vec{x}, \vec{y} \in \mathbb{R}^n$.

Corollary 38.6

Suppose we have a quadratic form $q(\vec{x}) : \mathbb{R}^n \to \mathbb{R}$, or quivalently its corresponding matrix $\mathbf{A}$.

We can say that q (equivalently $\mathbf{A}$) to be

- Positive definite if and only if all eigenvalues $\lambda_i > 0$.
- Negative definite if and only if all eigenvalues $\lambda_i < 0$.
- Positive semi-definite if and only if all eigenvalues $\lambda_i \geq 0$.
- Negative semi-definite if and only if all eigenvalues $\lambda_i \leq 0$.
- Indefinite if and only if some eigenvalues $\lambda_i > 0$ and some eigenvalues $\lambda_j > 0$.

Example 38.7

Let us look at $q(\vec{\boldsymbol{x}})_\beta = a_1^2 + 5a_3^2$ on $\mathbb{R}^3$, written in an orthonormal basis $\beta = \{\vec{\boldsymbol{v}}_1, \vec{\boldsymbol{v}}_2, \vec{\boldsymbol{v}}_3\}$.

We can see that q is positive semi-definite; plugging in $q(\vec{\boldsymbol{v}}_2) = q(\vec{\boldsymbol{e}}_2)_\beta = 0$.

38.3 Determining definiteness

In practice, with larger matrices and quadratic forms, it is infeasible to diagonalize $\mathbf{A}$ and find the eigenvalues. Can we bypass the diagonalization of $\mathbf{A}$ and still be able to decide if $\mathbf{A}$ is positive definite?

Yes, we can, using determinants!

We know that $\det \mathbf{A} = \lambda_1 \lambda_2 \dots \lambda_n$, and we want $\det \mathbf{A} > 0$ in order for $\mathbf{A}$ to be positive definite. However, this is only a necessary condition, but not a sufficient condition.

Definition 38.8: Principle Submatrix

The *principle submatrix* $\mathbf{A}^{(k)}$ of $\mathbf{A}$ is the square sub-matrix of size $k \times k$ starting at the top-left corner.

Example 38.9

Suppose we have

$$\mathbf{A} = \begin{bmatrix} 9 & -1 & 2 \\ -1 & 7 & -3 \\ 2 & -3 & 3 \end{bmatrix}.$$

We have

$$\begin{aligned} \det \mathbf{A}^{(3)} &= 89 > 0 \\ \det \mathbf{A}^{(2)} &= 62 > 0 \\ \det \mathbf{A}^{(1)} &= 9 > 0 \end{aligned}$$

Theorem 38.10: Positivity Criteria

A symmetric matrix $\mathbf{A}_{n\times n}$ is positive definite if and only if $\det \mathbf{A}^{(k)} > 0$ for all $k = 1, \dots, n$.

Let us look at $\mathbf{B}_{n\times m}$ over $\mathbb{R}$, and define $q(\vec{\boldsymbol{x}}) = \|\mathbf{B}\vec{\boldsymbol{x}}\|^2$ for $\vec{\boldsymbol{x}} \in \mathbb{R}^m$.

It turns out that this is a quadratic form, with matrix $\mathbf{A} = \mathbf{B}^T\mathbf{B}$, because

$$q(\vec{\boldsymbol{x}}) = \langle \mathbf{B}\vec{\boldsymbol{x}}, \mathbf{B}\vec{\boldsymbol{x}} \rangle = (\mathbf{B}\vec{\boldsymbol{x}})^T(\mathbf{B}\vec{\boldsymbol{x}}) = \vec{\boldsymbol{x}}^T\mathbf{B}^T\mathbf{B}\vec{\boldsymbol{x}}.$$

We already know that $\|\mathbf{B}\vec{\boldsymbol{x}}\|^2 \geq 0$, which means that $q(\vec{\boldsymbol{x}})$ is at least positive semi-definite.

We have that $q(\vec{\boldsymbol{x}})$ is positive definite if and only if $\mathbf{B}\vec{\boldsymbol{x}} = \vec{\mathbf{0}}$ only for $\vec{\boldsymbol{x}} = 0$, which happens if and only if $\operatorname{Ker} \mathbf{B} = \{\vec{\mathbf{0}}\}$, or $\mathbf{B}$ is one-to-one.

12/3/2021

Lecture 39

Gramian and Unitarily Equivalent

One problem we'll be focusing on today is: For which angles θ does there exist a basis for $\mathbb{R}^n$ such that $\angle(\vec{\boldsymbol{v}}_i, \vec{\boldsymbol{v}}_j) = \theta$, for $i \neq j$?

We already know that for 90°, we have orthonormal bases, but what about other angles?

Recall from last time that a quadratic form is a polynomial with only degree two terms. We can write this as $\vec{\boldsymbol{x}}^T\mathbf{A}\vec{\boldsymbol{x}}$, where $\mathbf{A}$ is a unique symmetric matrix. Diagonalizing, we have $[\vec{\boldsymbol{x}}]_\beta^T\mathbf{D}[\vec{\boldsymbol{x}}]_\beta$ in an orthonormal basis β. We further talked about positive definite quadratic forms, which satisfy $q(\vec{\boldsymbol{x}}) > 0$ for all $\vec{\boldsymbol{x}}$, or all eigenvalues $\lambda_i > 0$, or equivalently all principal submatrices $\det \mathbf{A}^{(k)} > 0$.

We ended with the fundamental example $q(\vec{\boldsymbol{x}}) = \|\mathbf{B}\vec{\boldsymbol{x}}\|^2$, where we can write this quadratic form as $\vec{\boldsymbol{x}}^T(\mathbf{B}^T\mathbf{B})\vec{\boldsymbol{x}}$, which is positive definite if and only if $\mathbf{B}$ is one-to-one, i.e. when $\operatorname{Ker} B = \{\vec{\mathbf{0}}\}$.

We can summarize this in a conclusion by noting that the set of quadratic forms (including zero) Q can be mapped to the set of symmetric matrices $\mathbf{A}$. It turns out that this map is linear, and is an isomorphism—the quadratic forms are isomorphic to symmetric matrices.

One question: why do we care so much about positive definite functions or matrices? It turns out that they define all inner products.

If we start with an inner product $\langle\cdot,\cdot\rangle$ on $\mathcal{V}^n$ over $\mathbb{R}$ or $\mathbb{C}$. We can use Gram-Schmidt to create an orthonormal basis β. We can express any inner product in this basis as $\langle \vec{\boldsymbol{x}}, \vec{\boldsymbol{y}} \rangle = [\vec{\boldsymbol{x}}]_\beta \circ [\vec{\boldsymbol{y}}]_\beta$; that is, any inner product is essentially just a standard dot product!

We can see why this is true by

$$\langle \vec{\boldsymbol{x}}, \vec{\boldsymbol{y}} \rangle = \left\langle \sum_i c_i \vec{\boldsymbol{v}}_i, \sum_j b_j \vec{\boldsymbol{v}}_j \right\rangle = \sum_{i,j} c_i \bar{b}_j \langle \vec{\boldsymbol{v}}_i, \vec{\boldsymbol{v}}_j \rangle = \sum_i c_i \bar{b}_i = [\vec{\boldsymbol{x}}]_\beta \circ [\vec{\boldsymbol{y}}]_\beta.$$

We can also create a quadratic form $q(\vec{\boldsymbol{x}}) = \langle \vec{\boldsymbol{x}}, \vec{\boldsymbol{x}} \rangle$; we already know that $\sum_{i=1}^n |c_i|^2 \geq 0$ always, with equality if and only if $\vec{\boldsymbol{x}} = 0$.

We can formally tell that this is a quadratic form by expanding $\vec{\boldsymbol{x}}$ as $\sum x_i \vec{\boldsymbol{e}}_i$, and expanding with linearity. We'd end up finding that $\mathbf{A}$ is symmetric.

Conversely, does any positive definite q give rise to an inner product?

Suppose we start with a positive definite quadratic form $q(\vec{\boldsymbol{x}}) : \mathbb{R}^n \to \mathbb{R}$. We need to construct a bilinear function $\langle \vec{\boldsymbol{x}}, \vec{\boldsymbol{y}} \rangle$ that reduces to $q(\vec{\boldsymbol{x}}) = \langle \vec{\boldsymbol{x}}, \vec{\boldsymbol{x}} \rangle$ for all $\vec{\boldsymbol{x}} = \vec{\boldsymbol{y}}$. If we further calculate $q(\vec{\boldsymbol{y}}) = \langle \vec{\boldsymbol{y}}, \vec{\boldsymbol{y}} \rangle$ and $q(\vec{\boldsymbol{x}} + \vec{\boldsymbol{y}}) = \langle \vec{\boldsymbol{x}} + \vec{\boldsymbol{y}}, \vec{\boldsymbol{x}} + \vec{\boldsymbol{y}} \rangle$, we can see that

$$\frac{1}{2}\big(q(\vec{\boldsymbol{x}} + \vec{\boldsymbol{y}}) - q(\vec{\boldsymbol{x}}) - q(\vec{\boldsymbol{y}})\big) = \langle \vec{\boldsymbol{x}}, \vec{\boldsymbol{y}} \rangle.$$

In conclusion, we have three sets that map to each other:

- Q^+, the set of all positive definite quadratic forms q
- S_n^*, the set of all positive definite symmetric matrices $\mathbf{A}$
- the set of all inner products on $\mathbb{R}^n$

Alternatively, we have the following three sets that map to each other:

- Q, the set of all quadratic forms q
- S_n, the set of all symmetric matrices $\mathbf{A}$
- the set of all bilinear forms on $\mathbb{R}^n$

Let's go back to the challenge problem. If we have a basis $\beta = \{\vec{v}_1, \ldots, \vec{v}_n\}$ of $\mathbb{R}^n$ such that $\angle(\vec{v}_i, \vec{v}_j) = \theta$ for all $i \neq j$, let us rescale all of them to make them unit; this does not change the angles.

We know that $\cos\theta = \frac{\vec{v}_i \circ \vec{v}_j}{\|\vec{v}_i\| \|\vec{v}_j\|}$. For $i \neq j$, we have $\vec{v}_i circ \vec{v}_j = \cos(\theta)$ because the denominator is 1, and $\vec{v}_i \circ \vec{v}_i = 1$ for all i.

As such, let us create a matrix $\mathbf{M} = (\vec{v}_i \circ \vec{v}_j)_{n \times n}$. We can see that this is a symmetric matrix, where 1's are along the diagonal, but $\cos\theta$ everywhere else.

We can rewrite this matrix as $\mathbf{B}^T\mathbf{B}$, where $\mathbf{B}$ is the matrix consisting of $\vec{v}i$'s the columns. This is because the dot products when we multiply the matrices are exactly the dot products we defined earlier.

In general,

Definition 39.1: Gramian

For any $\vec{v}_1, \ldots, \vec{v}_n \in \mathbb{R}^n$, we define their Gramian as

$$G(\vec{v}_1, \ldots, \vec{v}_n) = (\langle \vec{v}_i, \vec{v}_j \rangle)_{n \times n} = \mathbf{B}^T\mathbf{B}.$$

Here, $\mathbf{B}$ is the matrix consisting of $\vec{v}_i$'s as its columns.

How should we define this quadratic form to correspond to this Gramian? We'd need the matrix inside to be $\mathbf{B}^T\mathbf{B}$, giving us $q(\vec{x}) = \vec{x}^T(\mathbf{B}^T\mathbf{B})\vec{x}$.

We already know that this quadratic form is at least positive semi-definite, as it's the fundamental example we looked at earlier! Further, this quadratic is positive definite if and only if $\operatorname{Ker}\mathbf{B} = \{\vec{0}\}$, which we know to be true—our vectors form a basis for $\mathbb{R}^n$, and as such $\mathbf{B}$ must e invertible, and has a trivial kernel.

We've now solved our challenge problem: the only obstruction for θ is that the Gramian $\mathbf{M}_\theta$ is positive definite. But as a followup, for which θ is $\mathbf{M}_\theta$ positive definite?

Let us instead look at a more general case: when is the following matrix positive definite?

$$\mathbf{P} = \begin{bmatrix} a & & & \\ & a & b & \\ & b & \ddots & \\ & & & a \end{bmatrix}.$$

We found earlier that a matrix is positive definite if and only if the determinant of all principal submatrices are all positive.

We can see that $\det\mathbf{P} = (a-b)^{n-1} \cdot (a + (n-1)b)$, which we want to be > 0.

It turns out that $\mathbf{P}$ is positive if and only if $0 \leq b < a$ or $b < 0$ and $(n-1)(-b) < a$.

Back in our original problem, we have $a = 1$ and $b = \cos\theta$.

If $\cos\theta \geq 0$, we fall into the first case. We find that $\theta \leq 90°$, which corresponds to an acute or right angle. (That is, any acute angle works!)

If $\cos\theta < 0$, this corresponds to cases where $\theta > 90$, or obtuse angles. We need $(n-1)(-\cos\theta) < 1$, or $\cos\theta > -\frac{1}{n-1}$ in order for the Gramian to be positive definite.

It turns out that θ is restricted to

$$0 < \theta < \arccos\left(-\frac{1}{n-1}\right).$$

For $n = 2$, we have $\cos\theta > -\frac{1}{2-1} = -1$, which is always true. As such, all angles $0 < \theta < \pi$ are fine in the plane $\mathbb{R}^2$. Intuitively, this should make sense! Any two vectors (and thus any angle) forms a basis for $\mathbb{R}^2$.

For $n = 3$, we have $\cos\theta > -\frac{1}{3-1} = -\frac{1}{2}$, which is only true when $0 < \theta < 120°$.

For $n = 4$, we have $\cos\theta > -\frac{1}{4-1} = -\frac{1}{3}$, which is only true when $0 < \theta < \arccos\left(-\frac{1}{3}\right) \approx 109.5°$.

However, as we take the limit as $n \to \infty$, we have $\lim_{n\to\infty} \arccos\left(-\frac{1}{n-1}\right) = 90°$. As such, the only angles that work for *all* dimensions n are those such that $\theta \le 90$.

39.1 Spectral Theorem

Recall that the spectral theorem says that $\mathbf{A}$ over $\mathbb{C}$ is normal if ando nly if $\mathbf{A} = \mathbf{U}^{-1}\mathbf{D}\mathbf{U}$ where $\mathbf{U}$ is a unitary matrix, and $\mathbf{D}$ is a diagonal matrix.

For HW, we prove the backward direction; $\mathbf{A} = \mathbf{U}^{-1}\mathbf{D}\mathbf{U}$, which means that it us unitarily diagonalizable, and β is an orthonormal eigenbasis.

Suppose we look at

$$\begin{aligned} \mathbf{A}\mathbf{A}^* &= \left(\mathbf{U}^{-1}\mathbf{D}\mathbf{U}\right)\left(\mathbf{U}^{-1}\mathbf{D}\mathbf{U}\right)^* \\ &= \mathbf{U}^{-1}\mathbf{D}\mathbf{U} \cdot \mathbf{U}^*\mathbf{D}(\mathbf{U}^{-1})^* \\ &= \mathbf{U}^{-1}\mathbf{D}\mathbf{D}^*\mathbf{U} \end{aligned}$$

Flipping the two gives an equivalent result, crucially because diagonal matrices commute.

Definition 39.2: Unitarily/Orthogonally Equivalent

$\mathbf{A}$ and $\mathbf{B}$ are unitarily/orthogonally equivalent if

$$\mathbf{B} = \mathbf{U}\mathbf{A}\mathbf{U}^{-1}.$$

for some unitary $\mathbf{U}$ if and only if $\mathbf{A}$ is normal.