Vector Spaces

This week we will learn about:

- Abstract vector spaces,
- How to do linear algebra over fields other than $\mathbb{R}$,
- How to do linear algebra with things that don't look like vectors, and
- Linear combinations and linear (in)dependence (again).

Extra reading and watching:

- Sections 1.1.1 and 1.1.2 in the textbook
- Lecture videos 1, 1.5, 2, 3, and 4 on YouTube
- Vector space at Wikipedia
- Complex number at Wikipedia
- Linear independence at Wikipedia

Extra textbook problems:

⋆ 1.1.1, 1.1.4(a–f,h)

⋆⋆ 1.1.2, 1.1.5, 1.1.6, 1.1.8, 1.1.10, 1.1.17, 1.1.18

⋆⋆⋆ 1.1.9, 1.1.12, 1.1.19, 1.1.21, 1.1.22

☠ none this week

In the previous linear algebra course (MATH 2221), for the most part you learned how to perform computations with vectors and matrices. Some things that you learned how to compute include:

In this course, we will be working with many of these same objects, but we are going to generalize them and look at them in strange settings where we didn't know we could use them. For example:

In order to use our linear algebra tools in a more general setting, we need a proper definition that tells us what types of objects we can consider. The following definition makes this precise, and the intuition behind it is that the objects we work with should be "like" vectors in $\mathbb{R}^n$:

Definition 1.1 — Vector Space

Let $\mathcal{V}$ be a set and let $\mathbb{F}$ be a field. Let $\mathbf{v}, \mathbf{w} \in \mathcal{V}$ and $c \in \mathbb{F}$, and suppose we have defined two operations called *addition* and *scalar multiplication* on $\mathcal{V}$. We write the addition of $\mathbf{v}$ and $\mathbf{w}$ as $\mathbf{v} + \mathbf{w}$, and the scalar multiplication of c and $\mathbf{v}$ as $c\mathbf{v}$.

If the following ten conditions hold for all $\mathbf{v}, \mathbf{w}, \mathbf{x} \in \mathcal{V}$ and all $c, d \in \mathbb{F}$, then $\mathcal{V}$ is called a **vector space** and its elements are called **vectors**:

a) $\mathbf{v} + \mathbf{w} \in \mathcal{V}$ (closure under addition)

b) $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$ (commutativity)

c) $(\mathbf{v} + \mathbf{w}) + \mathbf{x} = \mathbf{v} + (\mathbf{w} + \mathbf{x})$ (associativity)

d) There exists a "zero vector" $\mathbf{0} \in \mathcal{V}$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$.

e) There exists a vector $-\mathbf{v}$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.

f) $c\mathbf{v} \in \mathcal{V}$ (closure under scalar multiplication)

g) $c(\mathbf{v} + \mathbf{w}) = c\mathbf{v} + c\mathbf{w}$ (distributivity)

h) $(c + d)\mathbf{v} = c\mathbf{v} + d\mathbf{v}$ (distributivity)

i) $c(d\mathbf{v}) = (cd)\mathbf{v}$

j) $1\mathbf{v} = \mathbf{v}$

Some points of interest are in order:

- A field $\mathbb{F}$ is basically just a set on which we can add, subtract, multiply, and divide according to the usual laws of arithmetic.

- Vectors might not look at all like what you're used to vectors looking like. Similarly, vector addition and scalar multiplication might look weird too (we will look at some examples).

Example. $\mathbb{R}^n$ *is a vector space.*

Example. $\mathcal{F}$*, the set of all functions* $f : \mathbb{R} \to \mathbb{R}$*, is a vector space.*

Example. *$\mathcal{M}_{m,n}(\mathbb{F})$, the set of all $m \times n$ matrices with entries from $\mathbb{F}$, is a vector space.*

Be careful: the operations that we call vector addition and scalar multiplication just have to satisfy the 10 axioms that were provided—they do not have to look *anything* like what we usually call "addition" or "multiplication."

Example. *Let $\mathcal{V} = \{x \in \mathbb{R} : x > 0\}$ be the set of positive real numbers. Define addition $\oplus$ on this set via usual multiplication of real numbers (i.e., $\mathbf{x} \oplus \mathbf{y} = xy$), and scalar multiplication $\odot$ on this set via exponentiation (i.e., $c \odot \mathbf{x} = x^c$). Show that this is a vector space.*

OK, so vectors and vector spaces can in fact look quite different from $\mathbb{R}^n$. However, doing math with them isn't much different at all: almost all facts that we proved in MATH 2221 actually only relied on the ten vector space properties provided a couple pages ago.

Thus we will see that really not much changes when we do linear algebra in this more general setting. We will re-introduce the core concepts again (e.g., subspaces and linear independence), but only very quickly, as they do not change significantly.

Complex Numbers

As mentioned earlier, the field $\mathbb{F}$ we will be working with throughout this course will always be $\mathbb{R}$ (the real numbers) or $\mathbb{C}$ (the complex numbers). Since complex numbers make linear algebra work so nicely, we give them a one-page introduction:

- We define i to be a number that satisfies $i^2 = -1$ (clearly, i is not a member of $\mathbb{R}$).
- An **imaginary number** is a number of the form bi, where $b \in \mathbb{R}$.
- A **complex number** is a number of the form $a + bi$, where $a, b \in \mathbb{R}$.
- Arithmetic with complex numbers works how you might naively expect:

  $$(a + bi) + (c + di) =$$

  $$(a + bi)(c + di) =$$

- Much like we think of $\mathbb{R}$ as a line, we can think of $\mathbb{C}$ as a plane, and the number $a + bi$ has coordinates (a, b) on that plane.
- The **length** (or **magnitude**) of the complex number $a + bi$ is $|a + bi| = \sqrt{a^2 + b^2}$.
- The **complex conjugate** of the complex number $a + bi$ is $\overline{a + bi} = a - bi$.
- We can use the previous facts to check that $(a + bi)\overline{(a + bi)} = |a + bi|^2$.
- We can also divide by (non-zero) complex numbers:

  $$\frac{a + bi}{c + di} =$$

Subspaces

It will often be useful for us to deal with vector spaces that are contained within other vector spaces. This situation comes up often enough that it gets its own name:

Definition 1.2 — Subspace

If $\mathcal{V}$ is a vector space and $\mathcal{S} \subseteq \mathcal{V}$, then $\mathcal{S}$ is a **subspace** of $\mathcal{V}$ if $\mathcal{S}$ is itself a vector space with the same addition and scalar multiplication as $\mathcal{V}$.

It turns out that checking whether or not something is a subspace is much simpler than checking whether or not it is a vector space. In particular, instead of checking all ten vector space axioms, you only have to check two:

Theorem 1.1 — Determining if a Set is a Subspace

Let $\mathcal{V}$ be a vector space and let $\mathcal{S} \subseteq \mathcal{V}$ be non-empty. Then $\mathcal{S}$ is a subspace of $\mathcal{V}$ if and only if the following two conditions hold for all $\mathbf{v}, \mathbf{w} \in \mathcal{S}$ and all $c \in \mathbb{F}$:

a) $\mathbf{v} + \mathbf{w} \in \mathcal{S}$ (closure under addition)

b) $c\mathbf{v} \in \mathcal{S}$ (closure under scalar multiplication)

Proof. For the "only if" direction,

For the "if" direction,

■

Example. *Is $\mathcal{P}^p$, the set of real-valued polynomials of degree at most p, a subspace of $\mathcal{F}$?*

Example. *Is the set of $n \times n$ real symmetric matrices a subspace of $\mathcal{M}_n(\mathbb{R})$?*

Example. *Is the set of 2×2 matrices with determinant 0 a subspace of $\mathcal{M}_2$?*

Spans, Linear Combinations, and Independence

We now present some definitions that you likely saw (restricted to $\mathbb{R}^n$) in your first linear algebra course. All of the theorems and proofs involving these definitions carry over just fine when replacing $\mathbb{R}^n$ by a general vector space $\mathcal{V}$.

Definition 1.3 — Linear Combinations

Let $\mathcal{V}$ be a vector space over the field $\mathbb{F}$, let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in \mathcal{V}$, and let $c_1, c_2, \ldots, c_k \in \mathbb{F}$. Then every vector of the form

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$$

is called a **linear combination** of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$.

Example. *Is a linear combination of and ?*

Example. *Is a linear combination of and ?*

Definition 1.4 — Span

Let $\mathcal{V}$ be a vector space and let $B \subseteq \mathcal{V}$ be a set of vectors. Then the **span** of B, denoted by $\mathrm{span}(B)$, is the set of all (finite!) linear combinations of vectors from B:

$$\mathrm{span}(B) \stackrel{\text{def}}{=} \left\{ \sum_{j=1}^{k} c_j \mathbf{v}_j \;\middle|\; k \in \mathbb{N},\ c_j \in \mathbb{F} \text{ and } \mathbf{v}_j \in B \text{ for all } 1 \leq j \leq k \right\}.$$

Furthermore, if $\mathrm{span}(B) = \mathcal{V}$ then $\mathcal{V}$ is said to be **spanned** by B.

Example. *Show that the polynomials* $1, x,$ *and* x^2 *span* $\mathcal{P}^2$.

Example. *Is* e^x *in the span of* $\{1, x, x^2, x^3, \ldots\}$?

Example. *Let $E_{i,j}$ be the matrix with a 1 in its (i,j)-entry and zeros elsewhere. Show that $\mathcal{M}_2$ is spanned by $E_{1,1}, E_{1,2}, E_{2,1}$, and $E_{2,2}$.*

Example. *Determine whether or not the polynomial $r(x) = x^2 - 3x - 4$ is in the span of the polynomials $p(x) = x^2 - x + 2$ and $q(x) = 2x^2 - 3x + 1$.*

Our primary reason for being interested in spans is that the span of a set of vectors is always a subspace (and in fact, we will see shortly that every subspace can be written as the span of some vectors).

Theorem 1.2 — Spans are Subspaces

Let $\mathcal{V}$ be a vector space and let $B \subseteq \mathcal{V}$. Then $\mathrm{span}(B)$ is a subspace of $\mathcal{V}$.

Proof. We just verify that the two defining properties of subspaces are satisfied:

■

Definition 1.5 — Linear Dependence and Independence

Let $\mathcal{V}$ be a vector space and let $B \subseteq \mathcal{V}$ be a set of vectors. Then B is **linearly dependent** if there exist scalars $c_1, c_2, \ldots, c_k \in \mathbb{F}$, at least one of which is not zero, and vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in B$ such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \mathbf{0}.$$

If B is not linearly dependent then it is called **linearly independent**.

There are a couple of different ways of looking at linear dependence and independence. For example:

- A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is linearly independent if and only if

  $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$ implies

- A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is linearly dependent if and only if there exists a particular j such that

  $\mathbf{v}_j$ is a

In particular, a set of two vectors is linearly dependent if and only if they are scalar multiples of each other.

Example. *Is the set of polynomials* $\Big\{ \qquad\qquad , \qquad\qquad \Big\}$ *linearly dependent or independent?*

Example. *Is the set of matrices* $\Bigg\{ \qquad\qquad , \qquad\qquad , \qquad\qquad \Bigg\}$ *linearly dependent or independent?*

Example. *Is the set of functions* $\{\sin^2(x), \cos^2(x), \cos(2x)\} \subset \mathcal{F}$ *linearly dependent or independent?*

Roughly, the reason that this final example didn't devolve into something we can just compute via "plug and chug" is that we don't have a nice basis for $\mathcal{F}$ that we can work with. This contrasts with the previous two examples (polynomials and matrices), where we do have nice bases, and we've been working with those nice bases already (perhaps without even realizing it).

We will talk about bases in depth next week!

Bases and Coordinate Systems

This week we will learn about:

- Bases of vector spaces,
- How to change bases in vector spaces, and
- Coordinate systems for representing vectors.

Extra reading and watching:

- Sections 1.1.3–1.2.2 in the textbook
- Lecture videos 5, 6, 7, and 8 on YouTube
- Basis (linear algebra) at Wikipedia
- Change of basis at Wikipedia

Extra textbook problems:

⋆ 1.1.3, 1.1.4(g), 1.2.1, 1.2.4(a–c,f,g)

⋆⋆ 1.1.15, 1.1.16, 1.2.2, 1.2.5, 1.2.7, 1.2.29

⋆⋆⋆ 1.1.17, 1.1.21, 1.2.9, 1.2.23

☠ 1.2.34

In introductory linear algebra, we learned a bit about bases, but we weren't really able to do too much with them when we were restricted to $\mathbb{R}^n$. Now that we are dealing with general vector spaces, bases will really start to shine, as they let us turn almost any vector space calculation into a familiar calculation in $\mathbb{R}^n$ (or $\mathbb{C}^n$).

Definition 2.1 — Bases

A **basis** of a vector space $\mathcal{V}$ is a set of vectors in $\mathcal{V}$ that

a) spans $\mathcal{V}$, and

b) is linearly independent.

Be careful: A vector space can have many bases that look very different from each other!

Example. *Let $\mathbf{e}_j$ be the vector in $\mathbb{R}^n$ with a 1 in its j-th entry and zeros elsewhere. Show that $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is a basis of $\mathbb{R}^n$.*

*[Side note: This is called the **standard basis** of $\mathbb{R}^n$.]*

Example. *Let $E_{i,j} \in \mathcal{M}_{m,n}$ be the matrix with a 1 in its (i,j)-entry and zeros elsewhere. Show that $\{E_{1,1}, E_{1,2}, \ldots, E_{m,n}\}$ is a basis of $\mathcal{M}_{m,n}$.*

*[Side note: This is called the **standard basis** of $\mathcal{M}_{m,n}$.]*

Example. *Show that the set of polynomials* $\{1, x, x^2, \ldots, x^p\}$ *is a basis of* $\mathcal{P}^p$.

[Side note: This is called the ***standard basis*** *of* $\mathcal{P}^p$.*]*

Example. *Is* $\{1 + x, 1 + x^2, x + x^2\}$ *a basis of* $\mathcal{P}^2$*?*

In the previous example, to answer a linear algebra question about $\mathcal{P}^2$, we converted the question into one about matrices, and then we answered that question instead. *This works in complete generality!* We will now start using bases to see that almost any linear algebra question that I can ask you about any vector space can be rephrased in terms of more "concrete" things like vectors in $\mathbb{R}^n$ and matrices in $\mathcal{M}_{m,n}$.

Our starting point is the following theorem:

Theorem 2.1 — Uniqueness of Linear Combinations

Let $\mathcal{V}$ be a vector space and let B be a basis for $\mathcal{V}$. Then for every $\mathbf{v} \in \mathcal{V}$, there is exactly one way to write $\mathbf{v}$ as a linear combination of the basis vectors in B.

Proof. The proof is very similar to the corresponding statement about bases of $\mathbb{R}^n$ from the previous course:

∎

The above theorem tells us that the following definition makes sense:

Definition 2.2 — Coordinate Vectors

Suppose $\mathcal{V}$ is a vector space over a field $\mathbb{F}$ with a finite (ordered) basis $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, and $\mathbf{v} \in \mathcal{V}$. Then the unique scalars c_1, c_2, …, $c_n \in \mathbb{F}$ for which

are called the **coordinates** of $\mathbf{v}$ with respect to B, and the vector

is called the **coordinate vector** of $\mathbf{v}$ with respect to B.

The above theorem and definition tell us that if we have a basis of a vector space, then we can treat the vectors in that space just like vectors in $\mathbb{F}^n$ (where n is the number of vectors in the basis). In particular, coordinate vectors respect vector addition and scalar multiplication "how you would expect them to:"

Example. *Find the coordinate vector of... ...with respect to the basis* $\{1, x, x^2\}$ *of* $\mathcal{P}^2$.

More generally,

Be careful: The order in which the basis vectors appear in B affects the order of the entries in the coordinate vector. This is kind of janky (technically, sets don't care about order), but everyone just sort of accepts it.

Example. *Find the coordinate vector of... ...with respect to the basis* $\{x^2, x, 1\}$ *of* $\mathcal{P}^2$.

Example. *Find the coordinate vector of... ...with respect to the basis* $\{1 + x, 1 + x^2, x + x^2\}$ *of* $\mathcal{P}^2$.

Notice that when we change the basis B that we are working with, coordinate vectors $[\mathbf{v}]_B$ change as well (even though $\mathbf{v}$ itself does not change). We will soon learn how to easily change coordinate vectors from one basis to another, but first we need to know that all coordinate vectors have the same number of entries:

Theorem 2.2 — Linearly Independent Sets versus Spanning Sets

Let $\mathcal{V}$ be a vector space with a basis B of size n. Then

a) Any set of more than n vectors in $\mathcal{V}$ must be linearly dependent, and

b) Any set of fewer than n vectors cannot span $\mathcal{V}$.

Proof. For (a), suppose there are $m > n$ vectors, which we call $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$. We want to solve $c_1\mathbf{v}_1 + \cdots + c_m\mathbf{v}_m = \mathbf{0}$. This is the same as

■

The previous theorem immediately implies the following one (which we proved for subspaces of $\mathbb{R}^n$ in the previous course):

Corollary 2.3 — Uniqueness of Size of Bases

Let $\mathcal{V}$ be a vector space that has a basis consisting of n vectors. Then *every* basis of $\mathcal{V}$ has exactly n vectors.

Based on the previous corollary, the following definition makes sense:

Definition 2.3 — Dimension of a Vector Space

A vector space $\mathcal{V}$ is called...

a) **finite-dimensional** if it has a finite basis, and its **dimension**, denoted by $\dim(\mathcal{V})$, is the number of vectors in one of its bases.

b) **infinite-dimensional** if it has no finite basis, and we say that $\dim(\mathcal{V}) = \infty$.

Example. *Let's compute the dimension of some vector spaces that we've been working with.*

Before proceeding, it is worth noting that every finite-dimensional vector space has a basis. The situation for infinite-dimensional vector spaces, however, is a bit murky...

Change of Basis

Sometimes one basis (i.e., coordinate system) will be much easier to work with than another. While it is true that the standard basis (of $\mathbb{R}^n$, $\mathbb{C}^n$, $\mathcal{P}^p$, or $\mathcal{M}_{m,n}$) is often the simplest one to use for calculations, other bases often reveal hidden structure that can make our lives easier.

We will discuss how to find these other bases shortly, but for now let's talk about how to convert coordinate systems from one basis to another.

Definition 2.4 — Change-of-Basis Matrix

Suppose $\mathcal{V}$ is a vector space with bases $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and C. The **change-of-basis matrix** from B to C, denoted by $P_{C \leftarrow B}$, is the $n \times n$ matrix whose columns are the coordinate vectors $[\mathbf{v}_1]_C, [\mathbf{v}_2]_C, \ldots, [\mathbf{v}_n]_C$:

The following theorem shows that the change-of-basis matrix $P_{C \leftarrow B}$ does exactly what its name suggests: it converts coordinate vectors from basis B to basis C.

Theorem 2.4 — Change-of-Basis Matrices

Suppose B and C are bases of a finite-dimensional vector space $\mathcal{V}$, and let $P_{C \leftarrow B}$ be the change-of-basis matrix from B to C. Then

a) $P_{C \leftarrow B}[\mathbf{v}]_B = [\mathbf{v}]_C$ for all $\mathbf{v} \in \mathcal{V}$, and

b) $P_{C \leftarrow B}$ is invertible and $P_{C \leftarrow B}^{-1} = P_{B \leftarrow C}$.

Furthermore, $P_{C \leftarrow B}$ is the unique matrix with property (a).

Some notes are in order:

Proof of Theorem 2.4. For (a), suppose $\mathbf{v} \in \mathcal{V}$ and write $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$, so that $[\mathbf{v}]_B = (c_1, c_2, \ldots, c_n)$. Then

■

Example. *Find the change-of-basis matrix* $P_{C \leftarrow B}$ *for the bases* $B = \{1, x, x^2\}$ *and* $C = \{1 + x, 1 + x^2, x + x^2\}$ *of* $\mathcal{P}^2$. *Then find the coordinate vector of...*
...with respect to C.

The previous example was not too difficult, since B happened to be the standard basis of $\mathcal{P}^2$. However, if it *weren't* the standard basis, then computing the columns of $P_{C\leftarrow B}$ would have been much more difficult (each column would require us to solve a linear system). The following theorem gives a better way of computing $P_{C\leftarrow B}$ in general:

Theorem 2.5 — Computing Change-of-Basis Matrices

Let $\mathcal{V}$ be a finite-dimensional vector space with bases B, C, and E. Then the reduced row echelon form of the augmented matrix

$$\left[\begin{array}{c|c} P_{E\leftarrow C} & P_{E\leftarrow B} \end{array}\right] \quad \text{is}$$

Proof. Suppose for now that we just wanted to compute $[\mathbf{v}_j]_C$ (the j-th column of $P_{C\leftarrow B}$).

∎

It is worth making some notes about the above theorem:

- $P_{E\leftarrow B}$ and $P_{E\leftarrow C}$ are both...

- This method for computing $P_{C \leftarrow B}$ is almost identical to the method you learned in introductory linear algebra for computing...

Example. *Find the change-of-basis matrix $P_{C \leftarrow B}$, where*

$$B = \Big\{ \quad , \quad , \quad \Big\} \quad \textit{and} \quad C = \Big\{ \quad , \quad , \quad \Big\}$$

are bases of . *Then compute* $[\mathbf{v}]_C$ *if* $[\mathbf{v}]_B = (1, 2, 3)$.

Linear Transformations

This week we will learn about:

- Linear transformations,
- The standard matrix of a linear transformation,
- Composition and powers of linear transformations, and
- Change of basis for linear transformations.

Extra reading and watching:

- Section 1.2.3 in the textbook
- Lecture videos 9, 10, 11, and 12 on YouTube
- Linear map at Wikipedia
- Transformation matrix at Wikipedia

Extra textbook problems:

⋆ 1.2.3

⋆⋆ 1.2.6, 1.2.11, 1.2.32

⋆⋆⋆ 1.2.12, 1.2.28, 1.2.30

☠ none this week

Last week, we learned that we could use bases to represent vectors in (finite-dimensional) vector spaces very concretely as tuples in $\mathbb{R}^n$ (or $\mathbb{F}^n$, where $\mathbb{F}$ is the field you're working in), thus turning almost any vector space problem into one that you learned how to solve in the previous course.

We will now introduce linear transformations between general vector spaces, and see that bases let us similarly think of any linear transformation (on finite-dimensional vector spaces) as a matrix in $\mathcal{M}_{m,n}$.

Definition 3.1 — Linear Transformations

Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces over the same field $\mathbb{F}$. A **linear transformation** is a function $T : \mathcal{V} \to \mathcal{W}$ that satisfies the following two properties:

a)

b)

Example. *Every matrix transformation is a linear transformation. That is,*

Example. *Is the function $T : \mathcal{M}_{m,n} \to \mathcal{M}_{n,m}$ that sends a matrix to its transpose a linear transformation?*

Example. *Is the function* $\det : \mathcal{M}_n \to \mathbb{R}$ *that sends a matrix to its determinant a linear transformation?*

Example. *Is the differentiation map* $D : \mathcal{D} \to \mathcal{F}$*, which sends a differentiable function to its derivative, a linear transformation?*

Before proceeding to prove things about linear transformations, we make some notes:

- We can sometimes consider the same linear transformation as acting on different vector spaces. For example, we can similarly consider D as a linear transformation from $\mathcal{P}^3$ to $\mathcal{P}^2$.
- For all linear transformations $T : \mathcal{V} \to \mathcal{W}$, it is true that $T(\mathbf{0}) = \mathbf{0}$.
- The **zero transformation** $O : \mathcal{V} \to \mathcal{W}$ is the one defined by
- The **identity transformation** $I : \mathcal{V} \to \mathcal{V}$ is the one defined by

The Standard Matrix

We now do for linear transformations what we did for vectors last week: we give them "coordinates" so that we can explicitly write them down using numbers in the ground field.

Theorem 3.1 — Standard Matrix of a Linear Transformation

Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces with bases B and D, respectively, where $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and $\mathcal{W}$ is m-dimensional. A function $T : \mathcal{V} \to \mathcal{W}$ is a linear transformation if and only if there exists a matrix $[T]_{D\leftarrow B} \in \mathcal{M}_{m,n}$ for which

Furthermore, the unique matrix $[T]_{D\leftarrow B}$ with this property is called the **standard matrix** of T with respect to the bases B and D, and it is

Before proving this theorem, we make some notes:

- The matrix $[T]_{D\leftarrow B}$ tells us how to convert coordinate vectors of $\mathbf{v} \in \mathcal{V}$ to coordinate vectors of $T(\mathbf{v}) \in \mathcal{W}$.
- Using this theorem, we can think of every linear transformation $T : \mathcal{V} \to \mathcal{W}$ as a matrix.
- The standard matrix looks different depending on the bases B and D,

Proof of Theorem 3.1. We just do block matrix multiplication:

∎

Standard matrices can perhaps be made a bit simpler to understand if we draw a schematic of how they work:

Example. *Find the standard matrix of the transpose map on $\mathcal{M}_2$ with respect to the standard basis $\{E_{1,1}, E_{1,2}, E_{2,1}, E_{2,2}\}$.*

Example. *Find the standard matrix of the differentiation map $D : \mathcal{P}^3 \to \mathcal{P}^3$ with respect to the standard basis $\{1, x, x^2, x^3\} \subset \mathcal{P}^3$.*

Composition and Powers of Linear Transformations

It is often useful to consider the effect of applying two or more linear transformations to a vector, one after another. Rather than thinking of these linear transformations as separate objects that are applied in sequence, we can combine their effect into a single new function that is called their **composition**:

The following theorem tells us that we can find the standard matrix of the composition of two linear transformations simply via matrix multiplication (as long as the bases "match up").

Theorem 3.2 — Composition of Linear Transformations

Suppose $\mathcal{V}$, $\mathcal{W}$, and $\mathcal{X}$ are finite-dimensional vector spaces with bases B, C, and D, respectively. If $T : \mathcal{V} \to \mathcal{W}$ and $S : \mathcal{W} \to \mathcal{X}$ are linear transformations then $S \circ T : \mathcal{V} \to \mathcal{X}$ is a linear transformation, and its standard matrix is

Proof. We just need to show that $[(S \circ T)(\mathbf{v})]_D = [S]_{D\leftarrow C}[T]_{C\leftarrow B}[\mathbf{v}]_B$ for all $\mathbf{v} \in \mathcal{V}$. To this end,

∎

In the special case when the linear transformations that we are composing are equal to each other, we get **powers** of those transformations:

In this special case, the previous theorem tells us that we can find the standard matrix of a power of a linear transformation by computing the corresponding power of the standard matrix of the original linear transformation.

Example. *Use standard matrices to compute the fourth derivative of* $x^2e^x + 2xe^x$.

Later on in this course, we will learn how to come up with a formula for powers of arbitrary matrices, which will let us (for example) find a formula for the n-th derivative of $x^2e^x + 2xe^x$.

Change of Basis for Linear Transformations

Recall that last week we learned how to convert a coordinate vector from one basis B to another basis C. We now learn how to do the same thing for linear transformations: we will see how to convert a standard matrix with respect to bases B and D to a standard matrix with respect to bases C and E.

Fortunately, we already did most of the hard work last week when we introduced change-of-basis matrices, so we can just "stitch things together" to make them work in this setting.

Theorem 3.3 — Change of Basis for Linear Transformations

Let $T : \mathcal{V} \to \mathcal{W}$ be a linear transformation between finite-dimensional vector spaces $\mathcal{V}$ and $\mathcal{W}$, and let B and C be bases of $\mathcal{V}$, while D and E are bases of $\mathcal{W}$. Then

The above theorem is made easier to remember by noting that adjacent subscripts always match (e.g., the two Ds are next to each other) and the outer subscripts on the left- and right-hand sides are the same (E's on the far left and C's on the far right).

We can also make sense of the theorem via a diagram:

Proof of Theorem 3.3. Let's think about what happens if we multiply $P_{E\leftarrow D}[T]_{D\leftarrow B}P_{B\leftarrow C}$ on the right by a coordinate vector $[\mathbf{v}]_C$:

■

Example. *Compute the standard matrix of the transpose map on $\mathcal{M}_2(\mathbb{C})$ with respect to the basis*

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \right\}.$$

Isomorphisms and Properties of Linear Transformations

This week we will learn about:

- Invertibility of linear transformations,
- Isomorphisms,
- Properties of linear transformations, and
- Non-integer powers of linear transformations.

Extra reading and watching:

- Sections 1.2.4 and 1.3.1 in the textbook
- Lecture videos 13, 14, 15, and 16 on YouTube
- Definition and Examples of Isomorphisms at WikiBooks
- Isomorphism at Wikipedia (be slightly careful – this page talks about isomorphisms on a broader context than just linear algebra)

Extra textbook problems:

⋆ 1.2.4(i,j), 1.3.1, 1.3.4(a–c), 1.3.5

⋆⋆ 1.2.10, 1.2.13–1.2.15, 1.2.17, 1.2.24, 1.2.25, 1.3.6

⋆⋆⋆ 1.2.19, 1.2.21, 1.2.33

☠ none this week

This week, we look at several important properties of linear transformations that you already saw for matrices back in introductory linear algebra. Thanks to standard matrices, all of these properties can be computed or determined using methods that we are already familiar with.

Invertibility of Linear Transformations

A linear transformation $T : \mathcal{V} \to \mathcal{W}$ is called **invertible** if there exists a linear transformation $T^{-1} : \mathcal{W} \to \mathcal{V}$ such that

The following theorem shows us that we can find the inverse of a linear transformation (if it exists) simply by inverting its standard matrix.

Theorem 4.1 — Invertibility of Linear Transformations

Let $T : \mathcal{V} \to \mathcal{W}$ be a linear transformation between n-dimensional vector spaces $\mathcal{V}$ and $\mathcal{W}$, which have bases B and D, respectively. Then T is invertible if and only if the matrix $[T]_{D\leftarrow B}$ is invertible. Furthermore,

$$([T]_{D\leftarrow B})^{-1} = \left[T^{-1}\right]_{B\leftarrow D}.$$

Proof. For the "only if" direction, note that if T is invertible then we have

■

Example. *Compute* $\int x^2 e^{3x}\,dx$. *(wait, what course is this?)*

Be careful: Differentiation is usually not an invertible transformation (why not?). The only reason it was invertible in the previous example was because we were able to choose the vector space $\mathcal{V}$ to not have any constant functions in it.

All of our methods of checking invertibility of matrices carry over straightforwardly to linear transformations on finite-dimensional vector spaces. For example...

Isomorphisms

Recall that every finite-dimensional vector space $\mathcal{V}$ has a basis B, and we can use that basis to represent a vector $\mathbf{v} \in \mathcal{V}$ as a coordinate vector $[\mathbf{v}]_B \in \mathbb{F}^n$, where $\mathbb{F}$ is the ground field. We used this correspondence between $\mathcal{V}$ and $\mathbb{F}^n$ to motivate the idea that...

We now make this idea of vector spaces being "the same" a bit more precise and clarify under exactly which conditions this "sameness" happens.

Definition 4.1 — Isomorphisms

Suppose $\mathcal{V}$ and $\mathcal{W}$ are vector spaces over the same field. We say that $\mathcal{V}$ and $\mathcal{W}$ are **isomorphic**, denoted by $\mathcal{V} \cong \mathcal{W}$, if there exists an invertible linear transformation $T : \mathcal{V} \to \mathcal{W}$ (called an **isomorphism** from $\mathcal{V}$ to $\mathcal{W}$).

The idea behind this definition is that if $\mathcal{V}$ and $\mathcal{W}$ are isomorphic then they have the same structure as each other—the only difference is the label given to their members ($\mathbf{v}$ for the members of $\mathcal{V}$ and $T(\mathbf{v})$ for the members of $\mathcal{W}$).

Example. *Show that $\mathcal{M}_{1,n}$ and $\mathcal{M}_{n,1}$ are isomorphic.*

Similarly, $\mathcal{M}_{1,n}$ and $\mathcal{M}_{n,1}$ are both isomorphic to...

Example. *Show that* $\mathcal{P}^3$ *and* $\mathbb{R}^4$ *are isomorphic.*

More generally, we have the following theorem that pins down the idea that every finite-dimensional vector space "behaves like" $\mathbb{F}^n$:

Theorem 4.2 — Isomorphisms of Finite-Dimensional Vector Spaces

Suppose $\mathcal{V}$ is an n-dimensional vector space over a field $\mathbb{F}$. Then $\mathcal{V} \cong \mathbb{F}^n$.

Proof. Pick some basis B of $\mathcal{V}$ and consider the function $T : \mathcal{V} \to \mathbb{F}^n$ defined by...

■

It is straightforward to check that if $\mathcal{V} \cong \mathcal{W}$ and $\mathcal{W} \cong \mathcal{X}$ then $\mathcal{V} \cong \mathcal{X}$. We thus get the following immediate corollary of the above theorem:

Properties of Linear Transformations

Now that we know we can think of arbitrary linear transformations (on finite-dimensional vector spaces) as matrices, we can apply all of our machinery from the previous course to them. For example, we can talk about the eigenvalues, range, null space, and rank of a linear transformation, and the definitions are just "what you would expect":

Furthermore, these properties can all be computed from the standard matrix.

Example. *Find the eigenvalues of the transposition map* $T : \mathcal{M}_2 \to \mathcal{M}_2$*, as well as a set of corresponding eigenvectors.*

Example. *Find the range and rank of the differentiation map $D : \mathcal{P}^3 \to \mathcal{P}^3$.*

Application: Diagonalization and Square Roots

Recall from introductory linear algebra that we can diagonalize many matrices. That is, for many $A \in \mathcal{M}_n$ we can write...

Doing so lets us easily take arbitrary (even non-integer) powers of matrices:

where D^r can simply be computed entrywise.

Thanks to standard matrices, we can now do the same thing for most linear transformations. We illustrate what we mean via an example.

Example. *Find a square root of the transpose map acting on* $\mathcal{M}_2$*.*

As perhaps an even more striking example, recall from last week that we could take powers of the standard matrix of the derivative to compute (for example) the fourth derivative of a function. If we use this method based on diagonalization to take non-integer powers of the standard matrix, we can compute *fractional* derivatives!

Example. *Compute the half-derivative of* $\sin(x)$ *and* $\cos(x)$*. Then find a formula for the* r*-th derivative of these functions for arbitrary (not necessarily integer)* $r \in \mathbb{R}$*.*

Inner Products and Orthogonality

This week we will learn about:

- Inner products (and the dot product again),
- The norm induced by the inner product,
- The Cauchy–Schwarz and triangle inequalities, and
- Orthogonality.

Extra reading and watching:

- Sections 1.3.4 and 1.4.1 in the textbook
- Lecture videos 17, 18, 19, 20, 21, and 22 on YouTube
- Inner product space at Wikipedia
- Cauchy–Schwarz inequality at Wikipedia
- Gram–Schmidt process at Wikipedia

Extra textbook problems:

⋆ 1.3.3, 1.3.4, 1.4.1

⋆⋆ 1.3.9, 1.3.10, 1.3.12, 1.3.13, 1.4.2, 1.4.5(a,d)

⋆⋆⋆ 1.3.11, 1.3.14, 1.3.15, 1.3.25, 1.4.16

☠ 1.3.18

There are many times when we would like to be able to talk about the angle between vectors in a vector space $\mathcal{V}$, and in particular orthogonality of vectors, just like we did in $\mathbb{R}^n$ in the previous course. This requires us to have a generalization of the dot product to arbitrary vector spaces.

Definition 5.1 — Inner Product

Suppose that $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, and $\mathcal{V}$ is a vector space over $\mathbb{F}$. Then an **inner product** on $\mathcal{V}$ is a function $\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \to \mathbb{F}$ such that the following three properties hold for all $c \in \mathbb{F}$ and all $\mathbf{v}, \mathbf{w}, \mathbf{x} \in \mathcal{V}$:

a) $\langle \mathbf{v}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{v} \rangle}$ (conjugate symmetry)

b) $\langle \mathbf{v}, \mathbf{w} + c\mathbf{x} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle + c\langle \mathbf{v}, \mathbf{x} \rangle$ (linearity in 2nd entry)

c) $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$, with equality if and only if $\mathbf{v} = \mathbf{0}$. (positive definiteness)

- Why those three properties?

- Inner products are *not* linear in their first argument...

- OK, so why does property (a) have that weird complex conjugation in it?

- For this reason, they are sometimes called "sesquilinear", which means...

Example. *Show that the following function is an inner product on* $\mathbb{C}^n$*:*

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^* \mathbf{w} = \sum_{i=1}^{n} \overline{v_i} w_i \quad \text{for all} \quad \mathbf{v}, \mathbf{w} \in \mathbb{C}^n.$$

Example. *Let* $a < b$ *be real numbers and let* $\mathcal{C}[a,b]$ *be the vector space of continuous functions on the interval* $[a,b]$*. Show that the following function is an inner product on* $\mathcal{C}[a,b]$*:*

$$\langle f, g \rangle = \int_a^b f(x) g(x) \; dx \quad \text{for all} \quad f, g \in \mathcal{C}[a,b].$$

The previous examples are the "standard" inner products on those vector spaces. However, inner products can also be much uglier. The following example illustrates how the same vector space can have multiple different inner products, and at first glance they might look nothing like the standard inner products.

Example. *Show that the following function is an inner product on* $\mathbb{R}^2$*:*

$$\langle \mathbf{v}, \mathbf{w} \rangle = v_1w_1 + 2v_1w_2 + 2v_2w_1 + 5v_2w_2 \quad \textit{for all} \quad \mathbf{v}, \mathbf{w} \in \mathbb{R}^2.$$

There is also a "standard" inner product on $\mathcal{M}_n$, but before being able to explain it, we need to introduce the following helper function:

Definition 5.2 — Trace

Let $A \in \mathcal{M}_n$ be a square matrix. Then the **trace** of A, denoted by $\mathrm{tr}(A)$, is the sum of its diagonal entries:

$$\mathrm{tr}(A) \stackrel{\text{def}}{=} a_{1,1} + a_{2,2} + \cdots + a_{n,n}.$$

Example. *Compute the following matrix traces:*

The reason why the trace is such a wonderful function is that it makes matrix multiplication "kind of" commutative:

Theorem 5.1 — Commutativity of the Trace

Let $A \in \mathcal{M}_{m,n}$ and $B \in \mathcal{M}_{n,m}$ be matrices. Then

$$\mathrm{tr}(AB) = \mathrm{tr}(BA).$$

Proof. Just directly compute the diagonal entries of AB and BA:

■

The trace also has some other nice properties that are easier to see:

With the trace in hand, we can now introduce the standard inner product on the vector space of matrices:

Example. *Show that the following function is an inner product on* $\mathcal{M}_{m,n}$*:*

$$\langle A, B \rangle = \mathrm{tr}(A^*B) \quad \textit{for all} \quad A, B \in \mathcal{M}_{m,n}.$$

The above inner product is typically called the **Frobenius inner product** or **Hilbert–Schmidt inner product**. Also, a vector space together with a particular inner product is called an **inner product space**.

Norm Induced by the Inner Product

Now that we have inner products, we can define the length of a vector in a manner completely analogous to how we did it with the dot product in $\mathbb{R}^n$. However, in this more general setting, we are a bit beyond the point of being able to draw a geometric picture of what length means (for example, what is the "length" of a continuous function?), so we change terminology slightly and instead call this function a "norm."

Definition 5.3 — Norm Induced by the Inner Product

Suppose that $\mathcal{V}$ is an inner product space. Then the **norm induced by the inner product** is the function $\|\cdot\| : \mathcal{V} \to \mathbb{R}$ defined by

$$\|\mathbf{v}\| \stackrel{\text{def}}{=} \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} \quad \text{for all} \quad \mathbf{v} \in \mathcal{V}.$$

Example. *What is the norm induced by the standard inner product on $\mathbb{C}^n$?*

Example. *What is the norm induced by the standard inner product on $\mathcal{C}[a, b]$?*

Example. *What is the norm induced by the standard (Frobenius) inner product on $\mathcal{M}_{m,n}$?*

Perhaps not surprisingly, the norm induced by an inner product satisfies the same basic properties as the length of a vector in $\mathbb{R}^n$. These properties are summarized in the following theorem.

Theorem 5.2 — Properties of the Norm Induced by the I.P.

Suppose that $\mathcal{V}$ is an inner product space, $\mathbf{v} \in \mathcal{V}$ is a vector, and $c \in \mathbb{F}$ is a scalar. Then the following properties of the norm induced by the inner product hold:

a) $\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$, and

b) $\|\mathbf{v}\| \geq 0$, with equality if and only if $\mathbf{v} = \mathbf{0}$.

The two other main theorems that we proved for the length in $\mathbb{R}^n$ were the Cauchy–Schwarz inequality and the triangle inequality. We now show that these same properties hold for the norm induced by any inner product.

Theorem 5.3 — Cauchy–Schwarz Inequality

Suppose that $\mathcal{V}$ is an inner product space and $\mathbf{v}, \mathbf{w} \in \mathcal{V}$. Then

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

Furthermore, equality holds if and only if $\{\mathbf{v}, \mathbf{w}\}$ is a linearly dependent set.

Proof. Let $c, d \in \mathbb{F}$ be arbitrary scalars, and expand $\|c\mathbf{v}+d\mathbf{w}\|^2$ in terms of the inner product:

■

For example, if we apply the Cauchy–Schwarz inequality to the Frobenius inner product on $\mathcal{M}_{m,n}$, it tells us that

and if we apply it to the standard inner product on $\mathcal{C}[a, b]$ then it says that

Neither of the above inequalities are particularly pleasant to prove directly.

Just as was the case in $\mathbb{R}^n$, the triangle inequality now follows very quickly from the Cauchy–Schwarz inequality.

Theorem 5.4 — The Triangle Inequality

Suppose that $\mathcal{V}$ is an inner product space and $\mathbf{v}, \mathbf{w} \in \mathcal{V}$. Then

$$\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|.$$

Furthermore, equality holds if and only if $\mathbf{v}$ and $\mathbf{w}$ point in the same direction (i.e., $\mathbf{v} = \mathbf{0}$ or $\mathbf{w} = c\mathbf{v}$ for some $0 \leq c \in \mathbb{R}$).

Proof. Start by expanding $\|\mathbf{v} + \mathbf{w}\|^2$ in terms of the inner product:

■

Orthogonality

The most useful thing that we can do with an inner product is re-introduce orthogonality in this more general setting:

Definition 5.4 — Orthogonality

Suppose $\mathcal{V}$ is an inner product space. Then two vectors $\mathbf{v}, \mathbf{w} \in \mathcal{V}$ are called **orthogonal** if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$.

In $\mathbb{R}^n$, we could think of "orthogonal" as a synonym for "perpendicular", since two vectors were orthogonal if and only if the angle between them was $\pi/2$. In general inner product spaces this geometric picture makes much less sense (for example, what does it mean for the angle between two polynomials to be $\pi/2$?), so it is perhaps better to think of orthogonal vectors as ones that are "as linearly independent as possible."'

With this intuition in mind, it is useful to extend orthogonality to *sets* of vectors, rather than just pairs of vectors:

Definition 5.5 — Orthonormal Bases

A basis B of an inner product space $\mathcal{V}$ is called an **orthonormal basis** of $\mathcal{V}$ if

a) $\langle \mathbf{v}, \mathbf{w} \rangle = 0$ for all $\mathbf{v} \neq \mathbf{w} \in B$, and (mutual orthogonality)

b) $\|\mathbf{v}\| = 1$ for all $\mathbf{v} \in B$. (normalization)

Example. *Examples of orthonormal bases in our "standard" vector spaces include...*

Orthogonal and orthonormal bases often greatly simplify calculations. For example, the following theorem shows us that linear independence comes for free when we know that a set of vectors are mutually orthogonal.

Theorem 5.5 — Orthogonality Implies Linear Independence

Let $\mathcal{V}$ be an inner product space and suppose that the set $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \subset \mathcal{V}$ consists of non-zero mutually orthogonal vectors (i.e., $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ whenever $i \neq j$). Then B is linearly independent.

Proof. Suppose $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$. Then...

■

A fairly quick consequence of the previous theorem is the fact that if a set of non-zero vectors is mutually orthogonal, and their number matches the dimension of the vector space, then...

Example. *Show that the set of Pauli matrices*

$$B = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \right\}$$

is an orthogonal basis of $\mathcal{M}_2(\mathbb{C})$*. How could you turn it into an orthonormal basis?*

We already learned that all finite-dimensional vector spaces are isomorphic (i.e., "essentially the same") to $\mathbb{F}^n$. It thus seems natural to ask the corresponding question about inner products—do all inner products on $\mathbb{F}^n$ look like the usual dot product on $\mathbb{F}^n$ in some basis? Orthonormal bases let us show that the answer is "yes."

Theorem 5.6 — All Inner Products Look Like the Dot Product

Suppose that B is an orthonormal basis of a finite-dimensional inner product space $\mathcal{V}$. Then

$$\langle \mathbf{v}, \mathbf{w} \rangle = [\mathbf{v}]_B \cdot [\mathbf{w}]_B \quad \text{for all} \quad \mathbf{v}, \mathbf{w} \in \mathcal{V}.$$

Proof. Write $B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_n\}$. Since B is a basis of $\mathcal{V}$, we can write $\mathbf{v} = c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n$ and $\mathbf{w} = d_1\mathbf{u}_1 + \cdots + d_n\mathbf{u}_n$. Then...

∎

If we specialize even further to $\mathbb{C}^n$ rather than to an arbitrary finite-dimensional vector space $\mathcal{V}$, then we can say even more. Specifically, recall that if $\mathbf{v}, \mathbf{w} \in \mathbb{C}^n$, E is the standard basis of $\mathbb{C}^n$, and B is any basis of $\mathbb{C}^n$, then

By plugging this fact into the above characterization of finite-dimensional inner product spaces (and assuming that B is orthonormal), we see that every inner product on $\mathbb{C}^n$ has the form

We state this fact in a slightly cleaner form below:

Corollary 5.7 — Invertible Matrices Make Inner Products

A function $\langle \cdot, \cdot \rangle : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ is an inner product if and only if there exists an invertible matrix $P \in \mathcal{M}_n(\mathbb{F})$ such that

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^*(P^*P)\mathbf{w} \quad \text{for all} \quad \mathbf{v}, \mathbf{w} \in \mathbb{F}^n.$$

For example, the usual inner product (i.e., the dot product) on $\mathbb{C}^n$ arises when $P = I$. Similarly, the weird inner product on $\mathbb{R}^2$ from a few pages ago, defined by

$$\langle \mathbf{v}, \mathbf{w} \rangle = v_1w_1 + 2v_1w_2 + 2v_2w_1 + 5v_2w_2 \quad \text{for all} \quad \mathbf{v}, \mathbf{w} \in \mathbb{R}^2,$$

is what we get if we choose $P = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$. To see this, we verify that

Orthogonalization

We already showed how to determine whether or not a particular set *is* an orthonormal basis, so let's turn to the question of how to *construct* an orthonormal basis. While this is reasonably intuitive in familiar inner product spaces like $\mathbb{R}^n$ or $\mathcal{M}_{m,n}(\mathbb{C})$, it becomes a bit more delicate when working in stranger inner products.

The process works one vector at a time to turn the vectors from some (not necessarily orthonormal) basis $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ into an orthonormal basis $C = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$. We start by simply defining

To construct the next member of our orthonormal basis, we define

In words, we are subtracting the portion of $\mathbf{v}_2$ that points in the direction of $\mathbf{u}_1$, leaving behind only the piece of it that is orthogonal to $\mathbf{u}_1$, as illustrated on the next page.

In higher dimensions, we would then continue in this way, adjusting each vector in the basis so that it is orthogonal to each of the previous vectors, and then normalizing it. The following theorem makes this precise and tells us that the result is indeed always an orthonormal basis.

Theorem 5.8 — Gram–Schmidt Process

Suppose $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis of an inner product space $\mathcal{V}$. Define

Then $C = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is an orthonormal basis of $\mathcal{V}$.

Proof. We actually prove that, not only is C an orthonormal basis of $\mathcal{V}$, but also that

for all $1 \leq k \leq n$.

We prove this result by induction on k. For the base case of $k = 1$, ...

■

Since finite-dimensional inner product spaces (by definition) have a basis consisting of finitely many vectors, and the Gram–Schmidt process tells us how to convert that basis into an orthonormal basis, we now know that every finite-dimensional inner product space has an orthonormal basis:

Corollary 5.9 — Existence of Orthonormal Bases

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

Example. *Find an orthonormal basis for* $\mathcal{P}^2[-1, 1]$ *with respect to the inner product*

$$\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)\ dx.$$

Advanced Linear Algebra – Week 6

Adjoints and Unitaries

This week we will learn about:

- The adjoint of a linear transformation, and
- Unitary transformations and matrices.

Extra reading and watching:

- Sections 1.4.2 and 1.4.3 in the textbook
- Lecture videos 23 and 24 on YouTube
- Unitary matrix at Wikipedia

Extra textbook problems:

⋆ 1.4.5(b,c,e,f), 1.4.8

⋆⋆ 1.4.3, 1.4.9–1.4.14, 1.4.21, 1.4.22

⋆⋆⋆ 1.4.6, 1.4.15, 1.4.18

☠ 1.4.19, 1.4.28

We now introduce the adjoint of a linear transformation, which we can think of as a way of generalizing the transpose of a real matrix to linear transformations between arbitrary inner product spaces.

Definition 6.1 — Adjoint Transformations

Suppose that $\mathcal{V}$ and $\mathcal{W}$ are inner product spaces and $T : \mathcal{V} \to \mathcal{W}$ is a linear transformation. Then a linear transformation $T^* : \mathcal{W} \to \mathcal{V}$ is called the **adjoint** of T if

For example, the adjoint of a matrix $A \in \mathcal{M}_{m,n}(\mathbb{R})$ is

Similarly, the adjoint of a matrix $A \in \mathcal{M}_{m,n}(\mathbb{C})$ is

So far, we have been a bit careless and referred to "the" adjoint of a matrix (linear transformation), even though it perhaps seems believable that a linear transformation might have more than one adjoint. The following theorem shows that, at least in finite dimensions, this is not actually a problem.

Theorem 6.1 — Existence and Uniqueness of Adjoints

Suppose that $\mathcal{V}$ and $\mathcal{W}$ are finite-dimensional inner product spaces. For every linear transformation $T : \mathcal{V} \to \mathcal{W}$ there exists a unique adjoint transformation $T^* : \mathcal{W} \to \mathcal{V}$. Furthermore, if B and C are orthonormal bases of $\mathcal{V}$ and $\mathcal{W}$ respectively, then

Proof. To prove uniqueness of T^*, suppose that T^* exists, let $\mathbf{v} \in \mathcal{V}$ and $\mathbf{w} \in \mathcal{W}$, and compute $\langle T(\mathbf{v}), \mathbf{w} \rangle$ in two different ways:

■

Example. *Show that the adjoint of the transposition map $T : \mathcal{M}_{m,n} \to \mathcal{M}_{n,m}$, with the Frobenius inner product, is also the transposition map.*

The situation presented in the above example, where a linear transformation is its own adjoint, is important enough that we give it a name:

Definition 6.2 — Self-Adjoint Transformations

Suppose that $\mathcal{V}$ is an inner product space. Then a linear transformation $T : \mathcal{V} \to \mathcal{V}$ is called **self-adjoint** if $T^* = T$.

For example, a matrix in $\mathcal{M}_n(\mathbb{R})$ is self-adjoint if and only if it is...

and a matrix in $\mathcal{M}_n(\mathbb{C})$ is self-adjoint if and only if it is...

Furthermore, a linear transformation is self-adjoint if and only if its standard matrix...

Unitary Transformations and Matrices

In situations where the norm of a vector is important, it is often desirable to work with linear transformations that do not alter that norm. We now start investigating these linear transformations.

Definition 6.3 — Unitary Transformations

Let $\mathcal{V}$ and $\mathcal{W}$ be inner product spaces and let $T : \mathcal{V} \to \mathcal{W}$ be a linear transformation. Then T is said to be **unitary** if

$$\|T(\mathbf{v})\| = \|\mathbf{v}\| \quad \text{for all} \quad \mathbf{v} \in \mathcal{V}.$$

We also say that a *matrix* is unitary if it acts as a unitary linear transformation on $\mathbb{F}^n$.

Example. *Show that the matrix* $U = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ *is unitary.*

Fortunately, there is a much simpler method of checking whether or not a matrix (or a linear transformations) is unitary, as demonstrated by the following theorem.

Theorem 6.2 — Characterization of Unitary Matrices

Suppose $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, and $U \in \mathcal{M}_n(\mathbb{F})$. The following are equivalent:

a) U is unitary,

b) $U^*U = I$,

c) $UU^* = I$,

d) $(U\mathbf{v}) \cdot (U\mathbf{w}) = \mathbf{v} \cdot \mathbf{w}$ for all $\mathbf{v}, \mathbf{w} \in \mathbb{F}^n$,

e) The columns of U are an orthonormal basis of $\mathbb{F}^n$, and

f) The rows of U are an orthonormal basis of $\mathbb{F}^n$.

It is worth comparing these properties to corresponding properties of invertible matrices:

Proof of Theorem 6.2. We do not prove all equivalences of this theorem – for that you can see the textbook. But we will demonstrate some of them in order to give an idea of why this theorem is true.

The equivalence of (b) and (c) follows from the fact that

To see that (d) $\implies$ (b), note that if we rearrange the equation $(U\mathbf{v}) \cdot (U\mathbf{w}) = \mathbf{v} \cdot \mathbf{w}$ slightly, we get

To see that (b) implies (a), suppose $U^*U = I$. Then for all $\mathbf{v} \in \mathbb{F}^n$ we have

To see that (b) is equivalent to (e), write U in terms of its columns $U = \big[\, \mathbf{u}_1 \mid \mathbf{u}_2 \mid \cdots \mid \mathbf{u}_n \,\big]$ and then use block matrix multiplication to multiply by U^*:

The remaining implications can be proved using similar techniques. ■

Checking whether or not a matrix is unitary is now quite simple, since we just have to check whether or not $U^*U = I$. For example, if we again return to the matrix

$$U = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

from earlier:

More generally, every rotation matrix and reflection matrix is unitary, as we now demonstrate.

Example. *Show that every rotation matrix $U \in \mathcal{M}_2(\mathbb{R})$ is unitary.*

Example. *Show that every reflection matrix $U \in \mathcal{M}_n(\mathbb{R})$ is unitary.*

In fact, the previous two examples provide exactly the intuition that you should have for unitary matrices—they are the ones that rotate and/or reflect $\mathbb{F}^n$, but do not stretch, shrink, or otherwise "distort" it. They can be thought of as "rigid" linear transformations that leave the size and shape of $\mathbb{F}^n$ in tact, but possibly change its orientation.

Schur Triangularization and the Spectral Decomposition(s)

This week we will learn about:

- Schur triangularization,
- The Cayley–Hamilton theorem,
- Normal matrices, and
- The real and complex spectral decompositions.

Extra reading and watching:

- Section 2.1 in the textbook
- Lecture videos 25, 26, 27, 28, and 29 on YouTube
- Schur decomposition at Wikipedia
- Normal matrix at Wikipedia
- Spectral theorem at Wikipedia

Extra textbook problems:

⋆ 2.1.1, 2.1.2, 2.1.5

⋆⋆ 2.1.3, 2.1.4, 2.1.6, 2.1.7, 2.1.9, 2.1.17, 2.1.19

⋆⋆⋆ 2.1.8, 2.1.11, 2.1.12, 2.1.18, 2.1.21

☠ 2.1.22, 2.1.26

We're now going to start looking at **matrix decompositions**, which are ways of writing down a matrix as a product of (hopefully simpler!) matrices. For example, we learned about diagonalization at the end of introductory linear algebra, which said that...

While diagonalization let us do great things with certain matrices, it also raises some new questions:

Over the next few weeks, we will thoroughly investigate these types of questions, starting with this one:

Schur Triangularization

We know that we cannot hope in general to get a diagonal matrix via unitary similarity (since not every matrix is diagonalizable via *any* similarity). However, the following theorem says that we can get partway there and always get an upper triangular matrix.

Theorem 7.1 — Schur Triangularization

Suppose $A \in \mathcal{M}_n(\mathbb{C})$. Then there exists a unitary matrix $U \in \mathcal{M}_n(\mathbb{C})$ and an upper triangular matrix $T \in \mathcal{M}_n(\mathbb{C})$ such that

Proof. We prove the result by induction on n (the size of A). For the base case, we simply notice that the result is trivial if $n = 1$: every 1×1 matrix is upper triangular.

∎

Let's make some notes about Schur triangularizations before proceeding...

- The diagonal entries of T are the eigenvalues of A. To see why, recall that the eigenvalues of a triangular matrix are its diagonal entries (theorem from previous course), and...

- The other pieces of Schur triangularization are

- To compute a Schur decomposition, follow the method given in the proof of the theorem:

The beauty of Schur triangularization is that it applies to *every* square matrix (unlike diagonalization), which makes it very useful when trying to prove theorems. For example...

Theorem 7.2 — Trace and Determinant in Terms of Eigenvalues

Suppose $A \in \mathcal{M}_n(\mathbb{C})$ has eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then

Proof. Use Schur triangularization to write $A = UTU^*$ with U unitary and T upper triangular. Then...

■

As another application of Schur triangularization, we prove an important result called the Cayley–Hamilton theorem, which says that every matrix satisfies its own characteristic polynomial.

Theorem 7.3 — Cayley–Hamilton

Suppose $A \in \mathcal{M}_n(\mathbb{C})$ has characteristic polynomial $p(\lambda) = \det(A - \lambda I)$. Then $p(A) = O$.

For example...

Proof of Theorem 7.3. Because we are working over $\mathbb{C}$, the Fundamental Theorem of Algebra says that we can factor the characteristic polynomial as a product of linear terms:

Well, let's Schur triangularize A:

$\blacksquare$

One useful feature of the Cayley–Hamilton theorem is that if $A \in \mathcal{M}_n(\mathbb{C})$ then it lets us write every power of A as a linear combination of $I, A, A^2, \ldots, A^{n-1}$. In particular,

Example. *Use the Cayley–Hamilton theorem to come up with a formula for A^4 as a linear combination of A and I, where*

$$A =$$

Example. *Use the Cayley–Hamilton theorem to find the inverse of the same matrix.*

Normal Matrices and the Spectral Decomposition

We now start looking at when Schur triangularization actually results in a *diagonal* matrix, rather than just an upper triangular one. We first need to introduce another new family of matrices:

Definition 7.1 — Normal Matrix

A matrix $A \in \mathcal{M}_n(\mathbb{C})$ is called **normal** if $A^*A = AA^*$.

Many of the important families of matrices that we are already familiar with are normal. For example...

However, there are also other matrices that are normal:

Example. *Show that the matrix* $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$ *is normal.*

Our primary interest in normal matrices comes from the following theorem, which says that normal matrices are exactly those that can be diagonalized by a unitary matrix:

Theorem 7.4 — Complex Spectral Decomposition

Suppose $A \in \mathcal{M}_n(\mathbb{C})$. Then there exists a unitary matrix $U \in \mathcal{M}_n(\mathbb{C})$ and diagonal matrix $D \in \mathcal{M}_n(\mathbb{C})$ such that

if and only if A is normal (i.e., $A^*A = AA^*$).

In other words, normal matrices are the ones with a diagonal Schur triangularization.

Proof. To see the "only if" direction, we just compute

∎

While we proved the spectral decomposition via Schur triangularization, that is not how it is computed in practice. Instead, we notice that the spectral decomposition is a special case of diagonalization where the invertible matrix that does the diagonalization is unitary, so we compute it via eigenvalues and eigenvectors (like we did for diagonalization last semester). Just be careful to choose the eigenvectors to have length 1 and be mutually orthogonal.

Example. *Find a spectral decomposition of the matrix...*

Example. *Find a spectral decomposition of the matrix*

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

Sometimes, we can just "eyeball" an orthonormal set of eigenvectors, but if we can't, we can instead apply the Gram–Schmidt process to any basis of the eigenspace.

The Real Spectral Decomposition

In the previous example, the spectral decomposition ended up making use only of real matrices. We now note that this happened because the original matrix was symmetric:

Theorem 7.5 — Real Spectral Decomposition

Suppose $A \in \mathcal{M}_n(\mathbb{R})$. Then there exists a unitary matrix $U \in \mathcal{M}_n(\mathbb{R})$ and diagonal matrix $D \in \mathcal{M}_n(\mathbb{R})$ such that

if and only if A is symmetric (i.e., $A^T = A$).

To give you a rough idea of why this is true, we note that every Hermitian (and thus every symmetric) matrix has real eigenvalues:

It follows that if A is Hermitian then we can choose the "D" piece of the spectral decomposition to be real. Also, it should not be too surprising, that if A is *real* and Hermitian (i.e., symmetric) that we can choose the "U" piece to be real as well.

We thus get the following 3 types of spectral decompositions for different types of matrices:

Geometrically, the real spectral decomposition says that real symmetric matrices are exactly those that act as follows:

Advanced Linear Algebra – Week 8

Positive (Semi)Definiteness

This week we will learn about:

- Positive definite and positive semidefinite matrices,
- Gershgorin discs and diagonal dominance,
- The principal square root of a matrix, and
- The polar decomposition.

Extra reading and watching:

- Section 2.2 in the textbook
- Lecture videos 30, 31, 32, and 33 on YouTube
- Positive-definite matrix at Wikipedia
- Gershgorin circle theorem at Wikipedia
- Square root of a matrix at Wikipedia
- Polar decomposition at Wikipedia

Extra textbook problems:

⋆ 2.2.1, 2.2.2

⋆⋆ 2.2.3, 2.2.5–2.2.10, 2.2.12

⋆⋆⋆ 2.2.11, 2.2.14, 2.2.16, 2.2.19, 2.2.22

☠ 2.2.18, 2.2.27, 2.2.28

Recall that normal matrices play a particularly important role in linear algebra (they can be diagonalized by unitary matrices). There is one particularly important family of normal matrices that we now focus our attention on.

Definition 8.1 — Positive (Semi)Definite Matrices

Suppose $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, and $A = A^* \in \mathcal{M}_n(\mathbb{F})$. Then A is called

a) **positive semidefinite (PSD)** if $\mathbf{v}^* A \mathbf{v} \geq 0$ for all $\mathbf{v} \in \mathbb{F}^n$, and

b) **positive definite (PD)** if $\mathbf{v}^* A \mathbf{v} > 0$ for all $\mathbf{v} \neq \mathbf{0}$.

Positive (semi)definiteness is somewhat difficult to eyeball from the entries of a matrix, and we should emphasize that it does *not* mean that the entries of the matrix need to be positive. For example, if

$$A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix},$$

then...

The definition of positive semidefinite matrices perhaps looks a bit odd at first glance. The next theorem characterizes these matrices in several other equivalent ways, some of which are hopefully a bit more illuminating and easier to work with.

Theorem 8.1 — Characterization of PSD and PD Matrices

Suppose $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, and $A = A^* \in \mathcal{M}_n(\mathbb{F})$. The following are equivalent:

a) A is positive (semidefinite | definite),

b) All of the eigenvalues of A are (non-negative | strictly positive),

c) There exists a diagonal $D \in \mathcal{M}_n(\mathbb{R})$ with (non-negative | strictly positive) diagonal entries and a unitary matrix $U \in \mathcal{M}_n(\mathbb{F})$ such that $A = UDU^*$, and

d) There exists (a matrix | an invertible matrix) $B \in \mathcal{M}_n(\mathbb{F})$ such that $A = B^*B$.

Proof. We prove the theorem by showing that (a) $\implies$ (b) $\implies$ (c) $\implies$ (d) $\implies$ (a).

∎

Example. *Show that* $A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ *is PSD, but not PD, in several different ways.*

Example. *Show that* $A = \begin{bmatrix} 2 & -1 & i \\ -1 & 2 & 1 \\ -i & 1 & 2 \end{bmatrix}$ *is positive definite.*

OK, let's look at another way of determining whether or not a matrix is positive definite, which has the advantage of not requiring us to compute eigenvalues.

Theorem 8.2 — Sylvester's Criterion

Let $A = A^* \in \mathcal{M}_n$. Then A is positive definite if and only if the determinant of the top-left $k \times k$ block of A is strictly positive for all $1 \leq k \leq n$.

We won't prove Sylvester's Criterion (a proof is in the textbook if you're curious), but instead let's jump right to an example to illustrate how it works.

Example. *Use Sylvester's criterion to show that* $A = \begin{bmatrix} 2 & -1 & i \\ -1 & 2 & 1 \\ -i & 1 & 2 \end{bmatrix}$ *is positive definite.*

Let's wrap up this section by reminding ourselves of something that we already proved about positive definite matrices a few weeks ago:

Theorem 8.3 — Positive Definite Matrices Make Inner Products

A function $\langle \cdot, \cdot \rangle : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ is an inner product if and only if there exists a positive definite matrix $A \in \mathcal{M}_n(\mathbb{F})$ such that

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^* A \mathbf{w} \quad \text{for all} \quad \mathbf{v}, \mathbf{w} \in \mathbb{F}^n.$$

Diagonal Dominance and Gershgorin Discs

In order to motivate this next section, let's think a bit about what Sylvester's criterion says when the matrix A is 2×2.

Theorem 8.4 — Positive Definiteness for 2×2 Matrices

Let $a, d \in \mathbb{R}$, $b \in \mathbb{C}$, and suppose that $A = \begin{bmatrix} a & b \\ \bar{b} & d \end{bmatrix}$.

a) A is positive semidefinite if and only if $a, d \geq 0$ and $|b|^2 \leq ad$, and

b) A is positive definite if and only if $a > 0$ and $|b|^2 < ad$.

Indeed, case (b) is exactly Sylvester's criterion. For case (a)...

Example. *Show that* $A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ *is positive semidefinite, but not positive definite.*

The previous theorem basically says that a 2×2 matrix is positive (semi)definite as long as its off-diagonal entries are "small enough" compared to its diagonal entries. This same intuition is well-founded even for larger matrices. However, to clarify exactly what we mean, we first need the following result that helps us bound the eigenvalues of a matrix based on simple information about its entries.

Theorem 8.5 — Gershgorin Disc Theorem

Let $A \in \mathcal{M}_n(\mathbb{C})$ and define the following objects:

- $r_i = \sum_{j \neq i} |a_{i,j}|$ (the sum of the off-diagonal entries of the i-th row of A),
- $D(a_{i,i}, r_i)$ is the closed disc in the complex plane centered at $a_{i,i}$ with radius r_i.

Then every eigenvalue of A is in at least one of the $D(a_{i,i}, r_i)$ (called **Gershgorin discs**).

Example. *Draw the Gershgorin discs for...*

Proof of Theorem 8.5. Let λ be an eigenvalue of A with associated eigenvector $\mathbf{v}$. Then...

∎

The Gershgorin disc theorem is an approximation theorem. For diagonal matrices we have $r_i = 0$ for all i, so the Gershgorin discs have radius 0 and thus the eigenvalues are exactly the diagonal entries (which we already knew from the previous course). However, as the off-diagonal entries increase, the radii of the Gershgorin discs increase so the eigenvalues can wiggle around a bit.

In order to connect Gershgorin discs to positive semidefiniteness, we introduce one additional family of matrices:

Definition 8.2 — Diagonally Dominant Matrices

Suppose that $A \in \mathcal{M}_n(\mathbb{C})$. Then A is called

a) diagonally dominant if $|a_{i,i}| \geq \sum_{j \neq i} |a_{i,j}|$ for all $1 \leq i \leq n$, and

b) strictly diagonally dominant if $|a_{i,i}| > \sum_{j \neq i} |a_{i,j}|$ for all $1 \leq i \leq n$.

Example. *Show that the matrix*

$$A = \begin{bmatrix} 2 & 0 & i \\ 0 & 3 & 1 \\ -i & 1 & 5 \end{bmatrix}$$

is strictly diagonally dominant, and draw its Gershgorin discs.

In particular, since the eigenvalues of the previous matrix were positive, it was necessarily positive definite. This same type of argument works in general, and leads immediately to the following theorem:

Theorem 8.6 — Diagonal Dominance Implies PSD

Suppose that $A = A^* \in \mathcal{M}_n(\mathbb{C})$ has non-negative diagonal entries.

a) If A is diagonally dominant then it is positive semidefinite.

b) If A is strictly diagonally dominant then it is positive definite.

Be careful: this is a one-way theorem! DD implies PSD, but PSD does not imply DD. For example,

Unitary Freedom of PSD Decompositions

We saw earlier that for every positive semidefinite matrix A we can find a matrix B such that $A = B^*B$. However, this matrix B is not unique, since if U is a unitary matrix and we define $C = UB$ then

The following theorem says that we can find *all* decompositions of A using this same procedure:

Theorem 8.7 — Unitary Freedom of PSD Decompositions

Suppose $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, and $B, C \in \mathcal{M}_{m,n}(\mathbb{F})$. Then $B^*B = C^*C$ if and only if there exists a unitary matrix $U \in \mathcal{M}_m(\mathbb{F})$ such that $C = UB$.

For the purpose of saving time, we do not show the “only if” direction of the proof here (it is in the textbook, in case you are interested).

The previous theorem raises the question of how simple we can make the matrix B in a positive semidefinite decomposition $A = B^*B$. The following theorem provides one possible answer: we can choose B so that it is also positive semidefinite.

Theorem 8.8 — Principal Square Root of a Matrix

Suppose $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, and $A \in \mathcal{M}_n(\mathbb{F})$ is positive semidefinite. Then there exists a unique positive semidefinite matrix $P \in \mathcal{M}_n(\mathbb{F})$, called the **principal square root** of A, such that

$$A = P^2$$

Proof. To see that such a matrix P exists, we use our usual diagonalization arguments.

■

The principal square root P of a matrix A is typically denoted by $P = \sqrt{A}$, and is in analogy with the principal square root of a non-negative real number (indeed, for 1×1 matrices they are the exact same thing).

Example. *Find the principal square root of...*

By combining our previous two theorems, we also recover a new matrix decomposition, which answers the question of how simple we can make a matrix by multiplying it on the left by a unitary matrix—we can always make it positive semidefinite.

Theorem 8.9 — Polar Decomposition

Suppose $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, and $A \in \mathcal{M}_n(\mathbb{F})$. Then there exists a unitary matrix $U \in \mathcal{M}_n(\mathbb{F})$ and a positive semidefinite matrix $P \in \mathcal{M}_n(\mathbb{F})$ such that

$$A = UP.$$

Proof. Since A^*A is positive semidefinite, we know from the previous theorem that

■

The matrix $\sqrt{A^*A}$ in the polar decomposition can be thought of as the "matrix version" of the absolute value of a complex number $|z| = \sqrt{\overline{z}z}$. In fact, this matrix is sometimes even denoted by $|A| = \sqrt{A^*A}$. Similarly, the polar decomposition of a matrix generalizes the polar form of a complex number:

We don't know how to compute the polar decomposition yet (since we skipped a proof earlier this week), but we will learn a method soon.

Over the past couple of weeks, we learned about several new families of matrices. It is worth drawing a diagram illustrating their relationships with each other:

It is also worth noting that many of these families of matrices are analogous to important subsets of the complex plane:

The Singular Value Decomposition

This week we will learn about:

- The singular value decomposition (SVD),
- Orthogonality of the fundamental matrix subspaces, and
- How the SVD relates to other matrix decompositions,

Extra reading and watching:

- Section 2.3.1 and 2.3.2 in the textbook
- Lecture videos 34, 35, 36, and 37 on YouTube
- Singular value decomposition at Wikipedia
- Fundamental Theorem of Linear Algebra at Wikipedia

Extra textbook problems:

⋆ 2.3.1, 2.3.4(a,b,c,f,g,i)

⋆⋆ 2.3.3, 2.3.5, 2.3.7

⋆⋆⋆ 2.3.14, 2.3.20

☠ 2.3.26

If the Schur decomposition theorem from last week was "big", then the upcoming theorem is "super-mega-gigantic". The singular value decomposition is possibly the biggest and most widely-used theorem in all of linear algebra (and is my personal favourite), so we're going to spend some time focusing on it.

Theorem 9.1 — Singular Value Decomposition (SVD)

Suppose $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, and $A \in \mathcal{M}_{m,n}(\mathbb{F})$. Then there exist unitary matrices $U \in \mathcal{M}_m(\mathbb{F})$ and $V \in \mathcal{M}_n(\mathbb{F})$ and a diagonal matrix $\Sigma \in \mathcal{M}_{m,n}(\mathbb{R})$ with non-negative entries such that

Furthermore, the diagonal entries of Σ (called the **singular values** of A) are the non-negative square roots of the eigenvalues of A^*A.

Let's compare how this decomposition theorem is good and bad compared to our previous decomposition theorems:

- Good:
- Good:
- Kinda good, kinda bad:

Proof. Consider the matrix A^*A and assume that $m \geq n$...

■

Some notes about the SVD are in order:

- The singular values of A are exactly the square roots of the eigenvalues of A^*A. Alternatively...

- Even though the singular values are uniquely determined by A, the diagonal matrix Σ isn't.

- The unitary matrices U and V are often not uniquely determined by A. Example:

Example. *Let's find the singular values of a matrix.*

To compute a full singular value decomposition (not just the singular values), we again leech off of diagonalization. Notice that

Similarly, the columns of U are eigenvectors of AA^*, but a slightly quicker (and slightly more correct) way to compute the columns of U is to notice that

Example. *Compute a singular value decomposition of the matrix*

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & 1 \\ 3 & 2 & 1 \end{bmatrix}.$$

Before delving into what makes the singular value decomposition so useful, it is worth noting that if $A \in \mathcal{M}_{m,n}(\mathbb{F})$ has singular value decomposition $A = U\Sigma V^*$ then A^T and A^* have singular value decompositions

In particular,

Geometric Interpretation

Recall that we think of unitary matrices as arbitrary-dimensional rotations and/or reflections. Using this intuition gives the singular value decomposition a simple geometric interpretation. Specifically, it says that every matrix $A = U\Sigma V^* \in \mathcal{M}_{m,n}(\mathbb{F})$ acts as a linear transformation from $\mathbb{F}^n$ to $\mathbb{F}^m$ in the following way:

- First,
- Then,
- Finally,

Let's illustrate this geometric interpretation in the $m = n = 2$ case:

In particular, it is worth keeping track not only of how the linear transformation changes a unit square grid on $\mathbb{R}^2$ into a parallelogram grid, but also how it transforms...

Furthermore, the two radii of the ellipse are exactly

In higher dimensions, linear transformations send (hyper-)ellipsoids to (hyper-)ellipsoids. For example, the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & 1 \\ 3 & 2 & 1 \end{bmatrix}$$

from earlier deforms the unit sphere as follows:

The fact that the unit sphere is turned into a 2D ellipse by this matrix corresponds to the fact that...

In fact, the first two left singular vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ (which point in the directions of the major and minor axes of the ellipse) form an orthonormal basis of the range.

This same type of argument works in general and leads to the following theorem:

Theorem 9.2 — Bases of the Fundamental Subspaces

Let $A \in \mathcal{M}_{m,n}$ be a matrix with rank r and singular value decomposition $A = U\Sigma V^*$, where

$$U = \left[\ \mathbf{u}_1 \mid \mathbf{u}_2 \mid \cdots \mid \mathbf{u}_m\ \right] \quad \text{and} \quad V = \left[\ \mathbf{v}_1 \mid \mathbf{v}_2 \mid \cdots \mid \mathbf{v}_n\ \right].$$

Then

a) $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ is an orthonormal basis of range(A),

b) $\{\mathbf{u}_{r+1}, \mathbf{u}_{r+2}, \ldots, \mathbf{u}_m\}$ is an orthonormal basis of null(A^*),

c) $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is an orthonormal basis of range(A^*), and

d) $\{\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \ldots, \mathbf{v}_n\}$ is an orthonormal basis of null(A).

Proof. Let's compute $A\mathbf{v}_j$:

■

Corollary 9.3 — Orthogonality of the Fundamental Subspaces

Suppose $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, and $A \in \mathcal{M}_{m,n}(\mathbb{F})$. Then

a) range(A) is orthogonal to null(A^*), and

b) null(A) is orthogonal to range(A^*).

In this corollary, when we say that one subspace is orthogonal to another, we mean that

Example. *Compute a singular value decomposition of the matrix*

$$A = \begin{bmatrix} 1 & 1 & 1 & -1 \\ 0 & 1 & 1 & 0 \\ -1 & 1 & 1 & 1 \end{bmatrix},$$

and use it to construct bases of the four fundamental subspaces of A.

Relationship With Other Matrix Decompositions

We now make sure that we really understand where the SVD fits into our world of matrix decompositions. For example, one way of rephrasing the singular value decomposition is as saying that we can always write a rank-r matrix as a sum of r rank-1 matrices in a very special way:

Theorem 9.4 — Orthogonal Rank-One Sum Decomposition

Suppose $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, and $A \in \mathcal{M}_{m,n}(\mathbb{F})$ is a matrix with $\text{rank}(A) = r$. Then there exist orthonormal sets of vectors $\{\mathbf{u}_i\}_{i=1}^r \subset \mathbb{F}^m$ and $\{\mathbf{v}_i\}_{i=1}^r \subset \mathbb{F}^n$ such that

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ are the non-zero singular values of A.

- This formulation is sometimes useful because...

- In fields other than $\mathbb{R}$ and $\mathbb{C}$, ...

Proof. For simplicity, we again assume that $m \leq n$ throughout this proof, and then we just do block matrix multiplication in the singular value decompositon:

■

In fact the singular value decomposition and the orthogonal rank-one sum decomposition are "equivalent" in the sense that you can prove one to quickly prove the other, and vice-versa. Sometimes they are both just called the singular value decomposition.

Example. *Compute an orthogonal rank-one sum decomposition of the matrix*

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

Similarly, the singular value decomposition is also "essentially equivalent" to the polar decomposition:

In the opposite direction,

If $A \in \mathcal{M}_n$ is positive semidefinite, then the singular value decomposition coincides exactly with the spectral decomposition:

A slight generalization of this type of argument leads to the following theorem:

Theorem 9.5 — Singular Values of Normal Matrices

Suppose $A \in \mathcal{M}_n$ is a normal matrix. Then the singular values of A are the absolute values of its eigenvalues.

Proof. Since A is normal, we can use the spectral decomposition to write $A = UDU^*$, where U is unitary and D is diagonal...

■

To see that the above theorem does not hold for non-normal matrices, consider the following example:

Example. *Compute the eigenvalues and singular values of the matrix*

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

Applications of the Singular Value Decomposition

This week we will learn about:

- The pseudoinverse of a matrix,
- The operator norm of a matrix, and
- Low-rank approximation and image compression.

Extra reading and watching:

- Section 2.3.3 and 2.C in the textbook
- Lecture videos 38, 39, 40, and 41 on YouTube
- Moore–Penrose inverse (pseudoinverse) at Wikipedia
- Operator norm at Wikipedia
- Low-rank approximation at Wikipedia

Extra textbook problems:

⋆ 2.3.2, 2.3.4(d,e,h), 2.C.4(a,b,d,e)

⋆⋆ 2.3.8–2.3.12, 2.C.1–2.C.3

⋆⋆⋆ 2.3.15, 2.3.21, 2.C.5, 2.C.6, 2.C.9, 2.C.10

☠ 2.3.17(a)

The Pseudoinverse

We have been working with the inverse of a matrix since early-on in introductory linear algebra, and while we can do great things with it, it has some deficiencies as well. For example, we know that if a matrix $A \in \mathcal{M}_n$ is invertible, then the linear system $A\mathbf{x} = \mathbf{b}$...

However, that linear system might have a solution even if A is *not* invertible. For example...

Example. *Show that the linear system*

$$\begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & 1 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \\ 6 \end{bmatrix}$$

has a solution, but its coefficient matrix is not invertible.

It thus seems natural to ask whether or not there exists a matrix $A^\dagger$ with the property that a solution to the linear system $A\mathbf{x} = \mathbf{b}$ (when it exists) is $\mathbf{x} = A^\dagger\mathbf{b}$. Well...

Definition 10.1 — Pseudoinverse of a Matrix

Suppose $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, and $A \in \mathcal{M}_{m,n}(\mathbb{F})$ has orthogonal rank-one sum decomposition

Then the **pseudoinverse** of A, denoted by $A^\dagger \in \mathcal{M}_{n,m}(\mathbb{F})$, is the matrix

There are several aspects of the pseudoinverse that we should clarify:

- If A is invertible, ...
- If A has SVD $A = U\Sigma V^*$, then...
- The pseudoinverse is well-defined.

Example. *Compute the pseudoinverse of the matrix*

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

The nice thing about the pseudoinverse is that it always exists (even if A is not invertible, or not even square), and it always finds a solution to the corresponding linear system (if a solution exists). Not only that, but if there are multiple different solutions, it finds the smallest one:

Theorem 10.1 — Pseudoinverses Solve Linear Systems

Suppose $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, $A \in \mathcal{M}_{m,n}(\mathbb{F})$, and suppose that the system of linear equations $A\mathbf{x} = \mathbf{b}$ has at least one solution. Then

is a solution. Furthermore, if $\mathbf{y}$ is any other solution then $\|A^\dagger \mathbf{b}\| < \|\mathbf{y}\|$.

Proof. We start by writing A in its orthogonal rank-one sum decomposition...

■

To get a rough idea for why it's desirable to find the solution with smallest norm, let's return to the linear system

$$\begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & 1 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \\ 6 \end{bmatrix}$$

from earlier. The solution set of this linear system consists of the vectors of the form

This solution set contains some vectors that are hideous, and some that are not so hideous:

The guarantee that the pseudoinverse finds the smallest-norm solution means that we do not have to worry about it returning "large and ugly" solutions like the first one above.

Geometrically, it means that the pseudoinverse finds the solution closest to the origin:

Not only does the pseudoinverse find the "best" solution when a solution exists, it even find the "best" non-solution when no solution exists!

This is strange to think about, but it makes sense if we again think in terms of norms and distances—if no solution to a linear system $A\mathbf{x} = \mathbf{b}$ exists, then it seems reasonable that the "next best thing" to a solution would be the vector that makes $A\mathbf{x}$ as close to $\mathbf{b}$ as possible. In other words, we want to find the vector $\mathbf{x}$ that...

The following theorem shows that choosing $\mathbf{x} = A^\dagger\mathbf{b}$ also solves this problem:

Theorem 10.2 — Linear Least Squares

Suppose $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, $A \in \mathcal{M}_{m,n}(\mathbb{F})$, and $\mathbf{b} \in \mathbb{F}^m$. If $\mathbf{x} = A^\dagger\mathbf{b}$ then

We won't prove this theorem (see the textbook if you're curious), but it comes up a lot in statistics, since it can be used to fit data to a model. For example, suppose we had 4 data points:

and we want to find a line of best fit for those data points (i.e., a line with the property that the sum of squares of vertical distances between the points and the line is minimized). To find this line, we consider the "ideal" scenario—we try (and typically fail) to find a line that passes exactly through all n data points by setting up the corresponding linear system:

Since this linear system has 4 equations, but only 2 variables (m and b), we do not expect to find an exact solution, but we can find the closest thing to a solution by using the pseudoinverse:

This exact same method also works for finding the "plane of best fit" for data points $(x_1, y_1, z_1), (x_2, y_2, z_2), \ldots, (x_n, y_n, z_n)$, and so on for higher-dimensional data as well. You can even do things like find quadratics of best fit, exponentials of best fit, or other weird functions of best fit.

By putting together all of the results of this section, we see that the pseudoinverse gives the "best solution" to a system of linear equations $A\mathbf{x} = \mathbf{b}$ in all cases:

The Operator Norm

We have seen one way of measuring the size of a matrix—the Frobenius norm. In practice, the Frobenius norm is actually not very useful (it's just used because it's easy to calculate), and the following norm is more commonly used instead:

Definition 10.2 — Operator Norm

Suppose $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, and $A \in \mathcal{M}_{m,n}(\mathbb{F})$. Then the **operator norm** of A, denoted by $\|A\|$, is either of the following (equivalent) quantities:

The operator norm is the maximum amount by which a matrix can stretch a vector:

Before showing that $\|A\|$ really is the largest singular value of A, let's establish some of its more basic properties.

Theorem 10.3 — Submultiplicativity

Suppose $A \in \mathcal{M}_{m,n}$ and $B \in \mathcal{M}_{n,p}$. Then

Proof. Notice that a matrix $A \in \mathcal{M}_{m,n}$ cannot stretch any vector by more than a factor of $\|A\|$:

■

Theorem 10.4 — Unitary Invariance

Let $A \in \mathcal{M}_{m,n}$ and suppose $U \in \mathcal{M}_m$ and $V \in \mathcal{M}_n$ are unitary matrices. Then

Proof. We start by showing that every unitary matrix $U \in \mathcal{M}_m$ has $\|U\| = 1$:

■

As a side note, the previous two theorems both hold for the Frobenius norm as well (try to prove these facts on your own). That is,

By combining unitary invariance with the singular value decomposition, we almost immediately confirm our observation that the operator norm should equal the matrix's largest singular value, and we also get a new formula for the Frobenius norm:

Theorem 10.5 — Matrix Norms in Terms of Singular Values

Let $A \in \mathcal{M}_{m,n}$ have rank r and singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. Then

Proof. If we write A in its singular value decomposition $A = U\Sigma V^*$, then unitary invariance tells us that $\|A\| = \|\Sigma\|$ and $\|A\|_{\mathrm{F}} = \|\Sigma\|_{\mathrm{F}}$. Well,

∎

Example. *Compute the operator and Frobenius norms of* $A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & 1 \\ 3 & 2 & 1 \end{bmatrix}$.

Low-Rank Approximation

As one final application of the singular value decomposition, we consider the problem of approximating a matrix by another matrix with small rank. One of the primary reasons why we would do this is that it allows us to compress data that is represented by a matrix, since a full $n \times n$ matrix requires us to store...

However, a rank-k matrix only requires us to store

Since $2kn$ is much smaller than n^2 when k is small, it is much less resource-intensive to store low-rank matrices than general matrices. Thus to compress data, instead of storing the exact matrix A that contains our data of interest, we can sometimes find a nearby matrix with small rank and store that instead.

To actually find a nearby low-rank matrix, we use the following theorem:

Theorem 10.6 — Eckart–Young–Mirsky

Suppose $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, and $A \in \mathcal{M}_{m,n}(\mathbb{F})$ has orthogonal rank-one sum decomposition

with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. Then the closest rank-$k$ matrix to A (i.e., the rank-k matrix B that minimizes $\|A - B\|$) is

In other words, the Eckart–Young–Mirsky theorem says that...

We skip the proof of the Eckart–Young–Mirsky theorem (see the textbook if you're curious), and instead jump right into a numerical example to illustrate its usage.

Example. *Find the closest rank-1 matrix to* $A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & 1 \\ 3 & 2 & 1 \end{bmatrix}$.

It is also worth noting that the Eckart–Young–Mirsky theorem works for many other matrix norms as well (like the Frobenius norm)—not just the operator norm.

One of the most interesting applications of this theorem is that it lets us do (lossy) image compression. We can represent an image by...

Applying the Eckart–Mirsky–Young theorem to those matrices then lets us compress the image. For example, let's use the following image:

Let's use MATLAB to compress the image by truncating its matrices' singular value decompositions: