MATH104 Notes

Adithya C. Ganesh

August 16, 2019

Contents

1   Lecture 4: 6-27-19

Last time, we defined a field, and a vector space. Today, we will define what a vector subspace is. Intuitively, it is a "smaller" vector space inside a bigger vector space.

We will also discuss properties of a vector space.

Recall that if you have a vector space V and a field $\mathbb{F}$, you can define two operations, vector addition and scalar multiplication as follows:

- $+ : V \times V \to V$
- $\times : \mathbb{F} \times V \to V$.

Definition. A vector space $W \subset V$ is a subspace of V if and only if for all $w_1, w_2 \in W$ and $\alpha, \beta \in \mathbb{F}$, we have

$$\alpha w_1 + \beta w_2 \in W.$$

Example. Note that $\mathbb{R}^2$ is a subspace of $\mathbb{R}^3$.

Example. Note that $\mathbb{R}^{n \times n}$ is a vector space over the field $\mathbb{R}$.

Let W be the space of all $n \times n$ symmetric matrices. Note that W is a subspace of $\mathbb{R}^{n \times n}$.

Let A, B be symmetric. Clearly, $\alpha A + \beta B$ is symmetric by linearity.

Example. Let $V = \mathbb{R}^{n \times n}, F = \mathbb{R}, W = \{A | A_{ij} = -A_{ji}\}$, the set of skew-symmetric matrices.

Let A, B be two matrices in W, and consider $\alpha, \beta \in \mathbb{R}$.

We would like to show that

$$(\alpha A + \beta B)_{ij} = -(\alpha A + \beta B)_{ji}.$$

And this follows from the same logic as before.

Example. Let's consider the vector space $V = \mathbb{R}^2$. Let $W = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} + c \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$.

So show that this isn't a subspace, let

$$w_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + c_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$w_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

And furthermore,

$$\alpha w_1 + \beta w_2 = (\alpha + \beta) \begin{pmatrix} 1 \\ 0 \end{pmatrix} + (\alpha c_1 + \beta c_2) \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

2

and we see that $aw_1 + bw_2 \notin W$. Furthermore, there isn't a zero vector in this subspace (if we look at it geometrically).

Definition. Let $v_1, \cdots, v_k$ be vectors. We say that the vectors are linearly dependent if there exists $\alpha_1, \ldots, a_k$ not all 0 such that

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k = 0.$$

We say that vectors are linearly independent if they are not linearly independent.

Example. What does it mean for a matrix to be linearly independent?

Let V be a matrix. We want to solve:

$$V = \begin{pmatrix} v_1 & v_2 & \ldots & v_k \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \ldots \\ \alpha_k \end{pmatrix} = 0$$

If the columns of V are linearly independent, then we can conclude that a is the zero vector.

If the columns are linearly dependent, you can find a nontrivial solution a to this problem.

Definition. We define the rank of a matrix V to be the maximum number of linearly independent columns.

Definition. We define the *span* of a set of vectors $v_1, \ldots, v_k$ to be the set

$$\{V | V = \alpha_1 v_1 + \cdots + \alpha_k v_k, \alpha_1, \ldots, a_k \in \mathbb{F}\}$$

Example. Let $V = \mathbb{R}^2$. What is the span of the vector $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$? You get $V = \begin{pmatrix} \alpha_1 \\ 0 \end{pmatrix}$. Geometrically, this is just the x-axis.

Example. What is the span of the vector $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$? You get $V = \begin{pmatrix} \alpha_1 \\ \alpha_1 \end{pmatrix}$, which is the linear $y = x$.

Example. What is the span of $\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}$? This is the whole plane.

Example. What is the span of $\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 \\ 0 \end{pmatrix}$? Just the x-axis.

Definition. Given a vector space V and a set of vectors $\{v_1, \ldots, v_n\}$, we say that the set is a basis if two conditions hold:

- $v_1, \ldots, v_n$ are linearly independent.
- span $\{v_1, \ldots, v_n\} = V$.

Example. The vectors $\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ are a basis of the plane $\mathbb{R}^2$.

Definition. The *dimension* of a vector space V is the number of vectors in a basis.

If $v_1, \ldots, v_n$ form a basis, then for any vector $v \in V$, we can choose constants $c_1, \ldots, c_n$ such that $V = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n$; note that these must be unique. Otherwise, we can get a contradiction by obtaining different constants $b_1, \cdots, b_n$ such that $V = b_1 v_1 + \cdots + b_n v_n$.

2 Lecture 5: 7-1-19

Recall the definition of linear dependence. If $v_1, \cdots, v_k$ are linearly dependent, there exists $c_1, \cdots, c_k$ not all 0 such that

$$c_1 v_1 + \cdots + c_k v_k = 0.$$

The vectors are linearly independent if the opposite is true.

Recall that the span of $v_1, \ldots, v_k$ is defined as follows:

$$\operatorname{span}\{v_1, \ldots, v_k\} = \{c_1 v_1 + \cdots + c_k v_k | c_i \in \mathbb{R}\}.$$

Recall that $v_1, \ldots, v_n$ is a basis for V if the v_i are linearly independent, and the v_i span V.

Theorem. If $v_1, \ldots, v_m$ form a basis, then any $v \in V$ can be expressed uniquely as

$$v = c_1 v_1 + \cdots + c_n v_n, c_i \in \mathbb{F}.$$

Definition. $c_1, c_2, \cdots, c_n$ are the coefficients of v with respect to the basis $v_1, \ldots, v_n$.

Example. Let $V = \mathbb{R}^3, \mathbb{F} = \mathbb{R}$. We can define the standard basis as follows:

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} ; \qquad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} ; \qquad \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} .$$

Then

$$v = 1\mathbf{e}_1 + 2\mathbf{e}_2 + 3\mathbf{e}_3 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} .$$

Now, consider a new basis

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} ; \qquad \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} ; \qquad \mathbf{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} .$$

The same vector v can be expressed as

$$v = \mathbf{v}_1 + \mathbf{v}_2 + 3\mathbf{v}_3.$$

Now we will talk a little bit about orthogonality, and sums and intersections of vector spaces.

Given two vectors $x, y \in \mathbb{R}^n$, we can define the inner product between x and y as follows:

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

If $x, y \in \mathbb{C}^n$, we define

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i \overline{y}_i = y^H x.$$

Definition. We say that x, y are orthogonal if $\langle x, y \rangle = 0$.

Definition. We say that $v_1, \ldots, v_n$ are mutually orthogonal if $\langle v_i, v_j \rangle = 0$ for any $i \neq j$.

Definition. We say that $v_1, \ldots, v_n$ are orthonormal if the vectors are mutually orthogonal and $\langle v_i, v_i \rangle = 1$.

Example. The standard basis in $\mathbb{R}^n$ is an orthonormal set of vectors.

Example. (Nonexample). The basis from before,

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}; \quad \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}; \quad \mathbf{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

is not an orthonormal basis, since

$$\langle v_1, v_1 \rangle = 2 \neq 1;$$
$$\langle v_1, v_2 \rangle = 1 \neq 0.$$

Definition. We say that a matrix A is orthogonal if $A^T A = I$. This is the same as requiring that the columns of A are orthonormal vectors.

Lemma. Suppose that $v_1, \ldots, v_n$ are mutually orthogonal in $\mathbb{R}^n$. Then $v_1, \ldots, v_n$ are linearly independent.

Proof. We want to show that if

$$c_1 v_1 + \cdots + c_n v_n = 0,$$

we must have $c_i = 0$ for all i.

Now, consider

$$\langle c_1 v_1 + \cdots + c_n v_n, v_i \rangle = \langle 0, v_i \rangle = 0.$$

Now, applying the linearity of the inner product, this is equal to:

$$c_1 \langle v_1, v_i \rangle + c_2 \langle v_2, v_i \rangle + \cdots + c_n \langle v_n, v_i \rangle = 0,$$

which is equal to $c_i \langle v_i, v_i \rangle = 0$. This implies that $c_i = 0$; and this is true for all i.

We will now talk about sums and intersections of subspaces.

Recall that if $W \subset V$, where V is a vector space over $\mathbb{F}$, we have W is a subspace of V iff for all $w_1, w_2 \in W$, $\alpha, \beta \in \mathbb{F}$, we have

$$\alpha w_1 + \beta w_2 \in W.$$

Suppose that $S, R \subset V$ are subspaces. We can define:

1. $S + R = \{s + r | s \in S, r \in R\}$.
2. $S \cap R = \{v | v \in S, v \in R\}$.

Lemma. $S + R$ and $S \cap R$ are subspaces.

Proof. Take $w_1, w_2 \in S+R$. By definition, there exists r_1, r_2, s_1, s_2 such that $w_1 = r_1 + s_1$, and $w_2 = r_2 + s_2$.

Now,

$$w_1 + w_2 = (s_1 + s_2) + (r_1 + r_2).$$

But $(s_1 + s_2) \in S$, and $(r_1 + r_2) \in R$, so $w_1 + w_2 \in S + R$ as claimed.

Similarly,

$$\alpha w_1 = \alpha(s_1 + r_1) = \alpha s_1 + \alpha r_1 \in S + R.$$

Exercise. Show that $S \cap R$ is a subspace.

Remark. Note that $S \cup R$ is not a subspace, the reason is that this isn't closed under addition. For example, the x-axis is a subspace of $\mathbb{R}^2$, and the y-axis is a subspace of $\mathbb{R}^2$. But the union doesn't contain vectors in the interior of the plane.

Definition. We say that $T = S \oplus R$, i.e. T is the direct sum of S and R.

1. We have $T = S + R$.
2. $S \cap R = \{0\}$.

Example. In $\mathbb{R}^3$, we can take

$$S = \text{span} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

$$R = \text{span} \left\{ \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

In this case, $\mathbb{R}^3 = S \oplus R$.

Example. Let $S_2 = \text{span} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $R_2 = \text{span} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$. Here, $S_2 \cap R_2 = \{0\}$, but $S_2 + R_2 \neq \mathbb{R}^3$.

Example. Let $S_3 = \text{span} \left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right)$, $R_3 = \text{span} \left(\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right)$. Herein, $S_3 + R_3 = \mathbb{R}^3$, but $S_3 \cap R_3 \neq \{0\}$.

Theorem. If $T = S \oplus R$, then every $t \in T$ can be expressed uniquely as $t = s + r$ for $s \in S, r \in R$.

Proof. Suppose for sake of contradiction, that $t = s_1 + r_1 = s_2 + r_2$. Then $s_1 - s_2 = r_2 - r_1$, but this is a contradiction since we assumed the intersection was just the zero vector.

Theorem. If $T = S \oplus R$, then $\dim T = \dim S + \dim R$.

Proof. The idea is that you can obtain a basis of T by choosing a basis of S, a basis of R, and combining them.

Theorem. We have $\dim S + \dim R = \dim S + \dim R - \dim S \cap R$. (Note that the direct sum theorem is a special case of this theorem.)

Note: no lecture on Thursday July 4.

3 Lecture 7: 7-3-19

Suppose V is a vector space over $\mathbb{R}$. Further, suppose V has dimension n, meaning that there exists a basis $v_1, \ldots, v_n$. Since this is a basis, for every $v \in V$, we can find constants $c_1, \ldots, c_n$ such that

$$V = c_1 v_1 + \cdots + c_n v_n.$$

Let $\varphi_B : V \to \mathbb{R}^n$ be a map such that

$$\varphi_B(v) = \begin{pmatrix} c_1 \\ \dots \\ c_n \end{pmatrix}.$$

Now, let $V = P^n = \{p | p(t) = a_0 + a_1 t + \dots + a_n t^n\}$. Here, we can choose our basis to be the monomials, that is

$$v_0 = 1$$
$$v_1 = t$$
$$v_2 = t^2$$
$$\dots$$
$$v_n = t^n.$$

Here, we can define a map $\varphi_B : P^n \to \mathbb{R}^{n+1}$, such that the map is defined as follows:

$$a_0 + a_1 t + \dots + a_n t^n \to \begin{pmatrix} a_0 \\ a_1 \\ \dots \\ a_n \end{pmatrix}.$$

A general fact. Given two polynomials, if

$$a_0 + a_1 t + \dots + a_n t^n = b_0 + b_1 t + \dots + b_n t^n,$$

for all t, then $b_i = a_i$ for all i.

A linear transformation is completely determined by where we send the basis vectors.

Suppose we have a map $L : V \to W$, where V, W are vector spaces over $\mathbb{R}$. Further, suppose $\dim V = n$ and $\dim W = m$.

- Let $v_1, \dots, v_n$ be a basis for V.

- Let $w_1, \dots, w_m$ be a basis for W.

Now, let $v = c_1 v_1 + \dots + c_n v_n$, and write

$$L(v) = L(c_1 v_1 + \dots + c_n v_n)$$
$$= c_1 L(v_1) + c_2 L(v_2) + \dots + c_n L(v_n).$$

Now, suppose

$$L(v_i) = a_1^i w_1 + a_2^i w_2 + \dots + a_m^i w_m$$

Matrix of linear map.

Let $L : V \to W$ be a linear map. We would like to find some A_L such that $L(v) = A_L v$. We can define the matrix as

$$A_L = \begin{pmatrix} \phi_{BC}(L(v_2)) & \phi_{BC}(L(v_2)) & \dots & \phi_{BC}(L(v_n)) \end{pmatrix}$$

Diagramatically, we can make a sort of square (commutative diagram):

$$L : V \to W$$
$$\phi_{BD} : V \to \mathbb{R}^n$$
$$A_L : R_n \to R_m$$
$$\phi_{BCD}^{-1} : \mathbb{R}^m \to W.$$

Last time, we discussed the derivative map. That is, given $p(t) = a_0 + a_1 t + \cdots + a_n t^n \in P^n$, we can define a map $L : P^n \to P^{n-1}$ known as the derivative map.

Note that we can compute the derivative matrix as follows:

$$A_L = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 2 & \ldots & 0 \\ \ldots & & & & \\ 0 & 0 & 0 & \ldots & n \end{pmatrix}$$

Change of basis matrix.

Let $v_1, \ldots, v_n$ be a basis for V. Let

$$B_1 = \begin{pmatrix} v_1 & v_2 & \ldots & v_n \end{pmatrix}.$$

And let

$$B_2 = \begin{pmatrix} w_1 & w_2 & \ldots & w_n \end{pmatrix}.$$

We would like to define a linear transformation that takes a vector in B_1 to B_2.

We get another diagram:

- $I_d : V \to V$
- $\phi_{B_2} : V \to R^n$
- $\phi_{B_1} : V \to R^n$.
- $M : \mathbb{R}^n \to \mathbb{R}^n$

Here, the change of basis M will take the coordinates in one basis and produce the coordinates in another basis.

If we have a linear transformation $L : V \to W$, we can compute the matrix $M_w A_L M_v^{-1}$, to obtain the change of basis matrix. Try to visualize this using the commutative diagram.

4 Lecture 8: 7-8-19

We start with continued discussion of the notion of change of basis matrices.

Suppose $S \in V$ is a subspace. Here, let

- V be a vector space over $\mathbb{R}$.
- We define $S^{\perp} := \{v \in V \,|\, \langle v, s \rangle = 0, \forall s \in S\}$.

For example, if $V = \mathbb{R}^2$, and $S = \text{span} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, then $S^{\perp} = \text{span} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. To prove this, we need to show $\text{span} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \subseteq S^{\perp}$ and $S^{\perp} \subseteq \text{span} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Properties of orthogonal complements.

- $S^\perp$ is a subspace of V.
- $S \oplus S^\perp = V$.
- $(S^\perp)^\perp = S$.
- $R \subseteq S$ is equivalent to $S^\perp \subseteq R^\perp$.
- $(R + S)^\perp = R^\perp \cap S^\perp$.
- $(S \cap R)^\perp = R^\perp + S^\perp$.

5   Lecture 9: 7-9-19

Recall that if $S \subset V$, where V is a vector space over $\mathbb{F}$, then

$$S^\perp = \{v \in V \,|\, v \perp s, s \in S\}.$$

The most important fact is that we can write

$$V = S \oplus S^\perp.$$

(Think about the geometry in $\mathbb{R}^n$.)

Definition. Given a linear map $A : V \to W$, we can define two relevant spaces.

- The kernel (or null space) of A is defined as follows:

$$\mathrm{Ker}\, A = \{v \in V \,|\, Av = 0\}.$$

- The image of A is defined as follows:

$$\mathrm{Im}\, A = \{Av \,|\, v \in V\}.$$

Example. Suppose that our linear transformation A is the projection onto e_1 (the x-axis), that is $\mathrm{Pr}_{e_1} : \mathbb{R}^2 \to \mathbb{R}^2$.

Then:

- $\mathrm{Ker}\, \mathrm{Pr}_{e_1} = $ the y axis.
- $\mathrm{Im}\, \mathrm{Pr}_{e_1} = $ the x axis.

We can also frame the kernel and the image in terms of the null-space and range. If $A : \mathbb{R}^n \to \mathbb{R}^n$ defines the map $v \mapsto Av$, then

- $\mathrm{Ker}\, A = \{v \in \mathbb{R}^n \,|\, Av = 0\} = N(A)$.
- $\mathrm{Im}\, A = \{Av \in \mathbb{R}^m \,|\, v \in \mathbb{R}^n\} = R(A)$.

For example, if our matrix is $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, then

- $\mathrm{Im}\, A = \mathrm{span}\begin{pmatrix} 1 \\ 0 \end{pmatrix}$.
- $\mathrm{Ker}\, A = \mathrm{span}\begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

Lemma. If $A = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n \end{pmatrix}$, then

$$\mathrm{Im}\, A = R(A) = \mathrm{span}\,\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}.$$

Proof. To prove the lemma, we will show containment in both directions. Suppose $v \in \text{span}\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$. Then for $c_i \in \mathbb{R}$, we have

$$v = c_1 a_1 + \cdots + c_n a_n$$

$$= \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n \end{pmatrix} \begin{pmatrix} c_1 \\ \ldots \\ c_n \end{pmatrix},$$

which implies that $v \in \text{Im } A$.

The opposite inclusion is equivalent; just reverse the steps. □

Now, given a matrix $A : V \to W$, we have four spaces associated with it. We have:

- $\text{Im}(A) \in W$
- $\text{Ker}(A) \in V$
- $\text{Im}(A^T) \in V$
- $\text{Ker}(A^T) \in W$.

Theorem. *Given a matrix A, we have the following important results relating the spaces associated with it.*

- $\text{Ker}(A) = \text{Im}(A^T)^{\perp}$
- $\text{Ker}(A^T) = \text{Im}(A)^{\perp}$.

Proof. We would like to show $\text{Ker}(A) = (\text{Im}(A^T))^{\perp}$. For all $v \in \text{Ker } A$, we want to show that v is orthogonal to all vectors in $\text{Im } A^T$.

Suppose that $x \in \text{Ker } A$. This implies $Ax = 0$. Also, for any $y \in \mathbb{R}^m$, we have $y^T Ax = 0$, but also $(A^T y)^T x = 0$. Further, $(A^T y)x = 0$ for all $y \in \mathbb{R}^m$.

This implies X is orthogonal to all vectors of the form $A^T y$ for all $y \in \mathbb{R}^m$. This means that $x \in \text{Im}(A^T)^{\perp}$.

For the opposite direction, suppose $x \in \text{Im}(A^T)^{\perp}$ for all $y \in \mathbb{R}^m$. Then $(A^T y)^T x = 0$, which implies $y^T Ax = 0$, that is $Ax = 0$, or $x \in \text{Ker } A$.

Hence, $\text{Im}(A^T)^{\perp} = \text{Ker } A$ as claimed. □

Corollary. In other words, $\text{Ker}(A)$ is perpendicular to $\text{Im}(A^T)$. This implies that $\text{Ker}(A) \oplus \text{Im}(A^T)^{\perp} = V$, so $\text{Ker}(A)$ and $\text{Im}(A^T)$ partition the space V.

6   Lecture 10: 7-10-19

Let A be a linear transformation taking $\mathbb{R}^n \to \mathbb{R}^m$. Then

$$\mathbb{R}^n = \text{Ker } A \oplus (\text{Ker } A)^{\perp} = \text{Ker } A \oplus \text{Im } A^T$$
$$\mathbb{R}^m = \text{Im } A \oplus \text{Ker } A^T.$$

Let $S \subset V$ be a subspace. If $V = S \oplus S^{\perp}$, then $\dim V = \dim S + \dim S^{\perp}$.

Theorem. *Let A be a map from $\mathbb{R}^n \to \mathbb{R}^m$. Further, suppose A_{res} denotes the map from $\text{Ker } A^{\perp} \to \Im A$. Then A_{res} is bijective.*

Proof. We start by showing that A_{res} is injective. Suppose $A_{res} v_1 = A_{res} v_2$. This implies $A_{res}(v_1 - v_2) = 0$. Since $v_1, v_2 \in \text{Ker } A^{\perp}$, this implies $v_1 - v_2 \in \text{Ker } A^{\perp}$ and $v_1 - v_2 \in \text{Ker } A$. In particular, this means that $v_1 - v_2 = 0$, so $v_1 = v_2$.

Now, we show that A_{res} is surjective. For all $w \in \operatorname{Im} A$, there exists some $v \in \mathbb{R}^n$ such that $Av = w$. Now, take $v_1 \in \operatorname{Ker} A$ and $v_2 \in \operatorname{Ker} A^\perp$ such that $v = v_1 + v_2$. Now, we know that $w = Av = A(v_1 + v_2) = Av_1 + Av_2 = Av_2 = w$, so $v_2 \in \operatorname{Ker} A^\perp$, and A_{res} is surjective.

This proves the desired result. $\qquad\qquad\square$

Definition. *We say that a map is a vector space isomorphism if it is bijective and a linear transformation.*

Theorem. *Suppose that $L : V \to W$ is a vector space isomorphism. Then:*

- *Given a basis $v_1, v_2, \ldots, v_n$ for V, then $L(v_1), L(v_2), \ldots, L(v_n)$ is a basis for W.*

- $\dim V = \dim W$.

Now, applying this result to the map between $\operatorname{Im} A \to \operatorname{Im} A^T$, we obtain the result that row rank is equal to column rank.

Also, remarking that $^n = \operatorname{Ker} A \oplus \operatorname{Im} A^T$, we can prove the rank nullity theorem (since dimensions add when you use direct sums).

Definition. *Suppose $f : S \to R$ is a function. We say that f is invertible if there exists $y : R \to S$ such that*

- $f(y(r)) = r; \quad r \in R$

- $g(f(s)) = s; \quad s \in S.$

Lemma. If f is bijective, then it is invertible.

Proof. For all $r \in R$, there exists a unique $s \in S$ such that $f(s) = r$. Define $g(r) = s$. $\qquad\square$

Now, suppose $L : V \to W$, and we take L^{-1}. Why is L^{-1} a linear map? Well, we can check it directly:

If

$$L^{-1}(\alpha w_1 + \beta w_2) = \alpha L^{-1}(w_1) + \beta L^{-1}(w_2),$$

then we can show that if we apply L to both sides, then we get equality, applying the linearity of L. Now, since $L(LHS) = L(RHS)$, we must have $LHS = RHS$, because L is injective.

Next lecture, we will discuss pseudoinverses. Given a linear transformation, if a linear transformation is invertible, we can write down the inverse pretty easily. What would happen if the linear transformation is not invertible? If we have a map from $\mathbb{R}^n \to \mathbb{R}^m$, we can redefine a notion of inverse that allows us to go back.

If $w = w_1 + w_2$, where $w_1 \in \operatorname{Im} A, w_2 \in \operatorname{Ker} A^T$, then $A^T w = A^\dagger w_1$, where $A^\dagger$ is known as the pseudoinverse. Importantly, the pseudoinverse of a matrix exists always.

Note that everything until tomorrow will be on the midterm (and everything starting on Monday will be on the final).

7 Lecture 11: 7-11-19

At this point, we're done with chapters 1, 2, and 3. We'll start discussing the notion of Moore-Penrose pseudoinverses.

Recall that $A \in \mathbb{R}^{n \times n}$ is invertible if there exists $B \in R^{n \times n}$ if

$$AB = BA = I.$$

Problem:

- There are matrices in $R^{n \times n}$ that are not invertible.

- The $n \times m$ matrices are not invertible.

Recall that if $A \in R^{n \times m}$, there are four fundamental spaces related to this matrix.

- $\text{Im}(A)$

- $\text{Ker}(A)$

- $\text{Im}(A^T)$

- $\text{Ker}(A^T)$.

Recall that we showed that:

- The orthogonal complement of the $\text{Im}\, A$ is $\text{Ker}(A^T)$; the orthogonal complement of $\text{Ker}(A)$ is $\text{Im}(A^T)$.

- We have the results:

$$\mathbb{R}^n = \text{Ker}\, A \oplus \text{Im}(A^T)$$
$$\mathbb{R}^m = \text{Ker}\, A^T \oplus \Im A.$$

- Also,

$$\text{Im}\, A^T \cong \text{Im}\, A.$$

Why is this useful? We have that $\mathbb{R}^n = \text{Ker}\, A \oplus \text{Im}\, A^\perp$.

This means that for every $v \in \mathbb{R}^n$, we can find a unique $v_1 \in \text{Ker}\, A$, $v_2 \in \text{Im}\, A^T$ such that $v = v_1 + v_2$.

Similarly, since $\mathbb{R}^m = \text{Ker}\, A^T \oplus \text{Im}\, A$, we can find a unique $w_1 \in \text{Ker}\, A^T$, $w_2 \in \text{Im}\, A$ such that $w = w_1 + w_2$.

Definition. Suppose we have a map $A : \mathbb{R}^n \to \mathbb{R}^m$; then the pseudoinverse is $A^+ : \mathbb{R}^m \to \mathbb{R}^n$. Further, suppose $w = w_1 + w_2$, where $w_1 \in \text{Ker}\, A^T$; $w_2 \in \text{Im}\, A$. We define:

$$A^+(w_1) = 0$$
$$A^+(w_2) = v_2,$$

where v_2 is the unique vector $\in \text{Im}\, A^T$ that satisfies $Av_2 = w_2$.

We can draw a picture explaining this:

Properties.

- If A is invertible, then $A^{-1} = A^+$.

- If B is the pseudoinverse of A, then

$$ABA = A.$$

Proof. Suppose we have some vector $v = v_1 + v_2$, with $v_1 \in \text{Ker}\, A$, $v_2 \in \text{Im}\, A^T$. Then

$$Av = A(v_1 + v_2)$$
$$= Av_2$$
$$= w_2.$$

Now, note that

$$BAv = Bw_2$$
$$= v_2.$$

Now,

$$ABAv = Av_2 = w_2.$$

$\square$

(Similarly, you can prove that $BAB = B$.)

- $(AB)^T = AB$.
- $(BA)^T = BA$.

In the homework, we will have to verify some of these properties.

Example. If $\alpha \in \mathbb{R}$, then

$$a^+ = \begin{cases} \frac{1}{a}; & \alpha \neq 0 \\ 0; & ; \alpha = 0. \end{cases}$$

Example. If A is surjective, then $\Im A$ is $\{0\}$, and

$$AA^+ = I_{m \times m}.$$

so A^+ is the right inverse.

In this setting, $A(v) = A(v_1 + v_2) = w_2$

Example. If A is injective, then $\mathrm{Ker}\, A = \{0\}$, and

$$A^+ A = I_{n \times n}.$$

Computationally, if you want to verify that a matrix is the pseudo inverse, it suffices to check the properties from before, i.e.

- $ABA = A$
- $BAB = B$
- $(AB)^T = AB$
- $(BA)^T = BA$.

Theorem. The pseudoinverse satisfies this limit:

$$\begin{aligned} A^+ &= \lim_{\delta \to 0} \left(A^T A + \delta^2 I\right)^{-1} A^T \\ &= \lim_{\delta \to 0} A^T (AA^T + \delta^2 I)^{-1}. \end{aligned}$$

Midterm will have 3 problems:

- Problem from homework.
- Problem from definition.
- Easy computational problem.

8   Lecture 12: 7-15-19

The last topic to appear on the midterm will be pseudoinverses. Recall HW4 is due Wednesday, and the exam is Thursday evening. Recall the structure of the midterm:

- Homework problem.

- Problem where you state a definition.

- Relatively similar computational problem.

Recall that if $A : \mathbb{R}^n \to \mathbb{R}^m$, we can decompose

$$\mathbb{R}^n = \operatorname{Ker} A \oplus \operatorname{Im} A^T$$
$$\mathbb{R}^m = \operatorname{Ker} A^T \oplus \operatorname{Im} A.$$

Recall the defintion of the pseudoinverse $A+ : \mathbb{R}^m \to \mathbb{R}^n$ such that for all $w \in \operatorname{Im} A$, we have

- $A^+(w) = v$ where v is the unique element in $\operatorname{Im} A^T$ such that $Av = w$.

- For all $w \in \operatorname{Ker} A^T$, we have $A^+(w) = 0$.

Also, B is the pseudoinverse of A iff

1. $ABA = A$

2. $BAB = B$

3. $(AB)^T = AB$

4. $(BA)^T = BA$.

The pseudoinverse limit definition is given by

$$A^+ = \lim_{\delta \to 0}(A^T A + \delta^2 I)^{-1}A^T$$
$$= \lim_{\delta \to 0} A(AA^T + \delta^2 I)^{-1}$$

Property 1. Recall that if we have a product UAV, where U and V are orthogonal, then

$$U^H U = UU^H = I,$$

and then

$$(UAV)^+ = V^H A^+ U^H.$$

Property 2. Recall that for a special class of diagonalizable matrices, we can express $A = UOU^H$, and thus $A^+ = UO^+U^H$.

The proofs of many of these results are provided in section 4.3.

Property 3. Recall that $(A^{-1})^T = (A^T)^{-1}$. We claim that

$$(A^+)^T = (A^T)^+.$$

To prove this, we will use the limit definition of the pseudoinverse.

Proof. Note that

$$
\begin{aligned}
(A^T)^+ &= \lim_{\delta \to 0} (AA^T + \delta^2 I)^{-1} A \\
&= \lim_{\delta \to 0} [A^T (A^T A + \delta^2 I)^{-1}]^T \\
&= [\lim_{\delta \to 0} A^T (A^T A + \delta^2 I)^{-1}]^T \\
&= (A^+)^T.
\end{aligned}
$$

$\square$

Recall the definition of eigenvectors, if $A \in \mathbb{C}^{n \times n}$, we know that $\mathbf{v}$ is a right eigenvector if there exists $\lambda \in \mathbb{C}$ such that

$$A\mathbf{v} = \lambda \mathbf{v}.$$

Similarly, $\mathbf{w}$ is a left eigenvector if there exists $\mu \in \mathbb{C}$ such that

$$w^H A = u w^H.$$

Recall that the eigenvalues of A are the roots of the characteristic polynomial $\det(A - \lambda I)$.

Also, note that for a 2×2 matrix, the characteristic polynomial is

$$\pi_A(\lambda) = \lambda^2 - \mathrm{Tr}(A)\lambda + \det A.$$

Suppose that $\pi_A(\lambda)$ is the characteristic polynomial of an $n \times n$ matrix, it has exactly n complex roots.

Definition. Note that we can write the characteristic polynomial as

$$\pi_A(\lambda) = (\lambda - \lambda_1)^{a_1} \ldots (\lambda - \lambda_m)^{a_m}.$$

The algebraic multiplicity of a root λ_i is a_i, where a_i is defined as in the equation above.

A natural question might be what is the geometric multiplicity? Based on these multiplicites, we get different Jordan canonical forms.

Definition. We can define the geometric multiplicity as

Geometric multiplicity is defined as

$$\dim \mathrm{Ker}(A - \lambda_i I)$$

In general, the geometric multiplicity is not equal to the algebraic multiplicity.

To compute the dimension of the kernel, use the rank nullity theorem.

Note that the algebraic multiplicity is $\geq$ the geometric multiplicity, and the geometric multiplicity is equal to the number of linearly independent eigenvectors.

Theorem. If for all λ_i eigenvalues of A, we have

$$AM(\lambda_i) = GM(\lambda_i),$$

then A is diagonalizable, and there exists V such that $V^H A V = D$, and thus $A = V D V^H$.

In this case, we have n linearly independent eigenvectors, and we can find a basis of $\mathbb{R}^n$ of eigenvectors. In other words, we can find a basis of $\mathbb{R}^n$ where the matrix A is diagonal with respect to the basis of eigenvectors.

9 Lecture 13: 7-16-19

Eastern european exams have a theoretical, oral component, and a practical component (problem solving). Homework 4 is due tomorrow at midnight.

Yesterday, given $A \in \mathbb{C}^{n \times n}$, we defined right eigenvectors in terms of the following equality:

$$Ax = \lambda x; \lambda \in \mathbb{C};$$

and similarly left eigenvectors satisfy

$$y^H A = \mu y^H.$$

- Step 1. Find the eigenvalues by finding the roots of the characteristic polynomials.
- Step 2. Solve the linear system in each eigenvalue to obtain the eigenvectors.

Recall that if $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then

$$A - \lambda I = \begin{pmatrix} a - \lambda & b \\ c & d - \lambda \end{pmatrix},$$

and

$$\det(A - \lambda I) = (a - \lambda)(d - \lambda) - bc;$$

and we can solve this system to obtain the eigenvalues.

In general, if A is $n \times n$

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

This is important because the product of the roots are equal to $\det A$, and the sum is equal to $\mathrm{Tr}(A)$. If the matrix A is singular, there exists $v \neq 0$ such that $Av = 0 = 0v$.

Example. Consider the matrix $A = \begin{pmatrix} 5 & -4 \\ 2 & -1 \end{pmatrix}$. Then

$$\begin{aligned} \det(A - \lambda I) &= (5 - \lambda)(-1 - \lambda) + 8 \\ &= (\lambda^2 - 4\lambda + 3) \\ &= (\lambda - 3)(\lambda - 1) = 0. \end{aligned}$$

So the eigenvalues are $\lambda = 1, 3$.

Now, to solve for the eigenvectors, we have to find a nontrivial element in the kernel of this matrix.

Definition. If $A \in \mathbb{C}^{n \times n}$, then the spectrum of A is defined as the set of eigenvalues.

Lemma. Eigenvalues are either real or come in conjugate pairs.

To see this, just note that λ is a root of the characteristic polynomial.

Also, note that all the eigenvectors are orthogonal (we will go over tomorrow).

Definition. Given $A \in \mathbb{R}^{n \times n}$, we say that A is diagonalizable if there exists an orthogonal Q, diagonal O such that

$$A = QOQ^H.$$

If A is not diagonalizable, we say that A is defective.

Theorem. If A has n distinct eigenvectors, then A is diagonalizable.

To see this, take D to be the diagonal matrix of eigenvalues, and Q to be the matrix whose columns are the eigenvectors.

Theorem. A is diagonalizable iff for all $\lambda_i \in \operatorname{Spec}(A)$, we have $GM(\lambda_i) = AM(\lambda_i)$.

10   Lecture 14: 7-17-19

Proposition. Suppose $v_1, \ldots, v_n$ are eigenvectors for $A \in \mathbb{R}^{n \times n}$ with different eigenvalues. Then, the eigenvectors are linearly independent.

Proof. We can prove this by induction. If $n = 2$, the result is clear, since if $\lambda_1 \neq \lambda_2$, we can write

$$Av_1 = \lambda_1 v_1$$
$$Av_2 = \lambda_2 v_2.$$

Assume for sake of contradiction $v_1 = cv_2$, where $c \neq 0$. Now, if $Av_1 = cAv_2 = cA_2 v_2$, this implies $A_1 v_1 = \lambda_1 c v_2$, which implies $\lambda_1 = \lambda_2$, which is a contradiction.

By induction, assume that $v_1, \ldots, v_j$ are linearly independent with distinct eigenvalues $\lambda_1, \ldots, \lambda_j$, for some $2 < j < n$. By contradiction, assume that for the next eigenvector v_{j+1}, $Av_{j+1} = \lambda_{j+1} v_{j+1}$, and $\lambda_{j+1} \neq \lambda_i, i \in \{1, \ldots, j\}$. Now, if

$$v_{j+1} = \sum_{i=1}^{j} c_i v_i,$$

we can write

$$\lambda_{j+1} \sum_{i=1}^{j} c_i v_i = Av_{j+1} = A \sum_{i=1}^{j} c_i v_i$$
$$= \sum_{i=1}^{j} c_i Av_i$$
$$= \sum_{i=1}^{j} c_i \lambda_i v_j.$$

Now, term by term, this is equivalent to saying that

$$\sum_{i=1}^{j} c_i (\lambda_i - \lambda_{j+1}) v_j = 0.$$

But since not all the c_i's are 0, this implies that the v_j's are linearly dependent, which is a contradiction. □

Theorem. Let A be a symmetric matrix, with $A^T = A$, with $A \in \mathbb{R}^{n \times n}$, and x, y are eigenvectors with different eigenvalues λ, μ. Then $\langle x, \rangle = 0$.

Proof. Note that

$$\langle Ax, y \rangle = (Ax)^T y$$
$$= x^T A^T y$$
$$= \langle x, A^T y \rangle.$$

Now,

$$\begin{aligned}
\lambda \langle x, y \rangle &= \langle \lambda x, y \rangle \\
&= \langle Ax, y \rangle \\
&= \langle x, A^T y \rangle \\
&= \langle x, Ay \rangle \\
&= \langle x, \mu y \rangle \\
&= \mu \langle x, y \rangle.
\end{aligned}$$

Since $\lambda \neq \mu$, it follows that $\langle x, y \rangle = 0$. $\qquad \square$

Theorem. Let $A \in \mathbb{R}^{n \times n}$. If x, y are right and left eigenvectors with $\lambda \neq \mu$, it turns out that $\langle x, y \rangle = 0$.

Proof. Note that

$$\begin{aligned}
\mu \langle x, y \rangle &= \langle \mu x, y \rangle \\
&= \langle Ax, y \rangle \\
&= \langle x, A^T y \rangle \\
&= \langle x, (y^T A)^T \rangle \\
&= \langle x, (\lambda y^T)^T \rangle \\
&= \langle x, \lambda a \rangle \\
&= \lambda \langle x, y \rangle.
\end{aligned}$$

$\qquad \square$

Theorem. If $A \in \mathbb{R}^{n \times n}$ matrix, then A and A^T have the same eigenvalues.

Proof. Clearly,

$$\begin{aligned}
\pi_A(\lambda) &= \det(A - \lambda I) \\
&= \det((A - \lambda I)^T) \\
&= \det(A^T - \lambda I) \\
&= \pi_{A^T}(\lambda).
\end{aligned}$$

$\qquad \square$

Corollary. The left and right eigenvalues of A are the same.

Theorem. Let $A \in \mathbb{C}^{n \times n}$ with $\lambda_1, \ldots, \lambda_n$ distinct and right eigenvectors $x_1, \ldots, x_n$, and left eigenvectors $y_1, \ldots, y_n$ such that $y_j^H x_i = \delta_{ij}$[1]

Let $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, $X = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix}$, $Y = \begin{pmatrix} y_1 & y_2 & \cdots & y_n \end{pmatrix}$. Then A is diagonalizable as $A = XD^{X-1} = XDY^H$.

Proof. Then:

- $AX = XD$
- $y^H A = DY^H$
- $y^H X = I$

---

[1] $\delta_{ij} = 1$ if $i = j$, 0 otherwise.

- $A = XDX^{-1} = XDY^H = \sum_{i=1}^{n} \lambda_i x_i y_i^H.$

$\square$

Theorem. (Spectral theorem). If $A \in \mathbb{R}^{n \times n}$ and $A^T = A$, then A has n eigenvectors that are orthonormal. Then

- $AQ = QD$
- $A = QDQ^T \sum_{i=1}^{n} \lambda_i \mu_i \mu_i^H.$

The only thing we didn't show is that n eigenvectors exist. Once we get this result, we can show the fact that A is diagonalizable.

This turns out to be computationally useful in a lot of cases.

For next time, we will continue with similar matrices and matrix exponential.

11 Lecture 15: 7-18-19

Definition. We say that a matrix A is similar to matrix B if

$$A = SBS^{-1},$$

for some invertible matrix S.

Example. If A has n linearly independent eigenvectors, then

$$AX = XD,$$

where $X = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix}$, and $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. In this case, $A = XDX^{-1}$.

Properties of similar matrices.

- If A and B are similar, then $\det(A) = \det(B)$. This follows directly from the multiplicative property of the determinant.

  Proof. $\det(A) = \det(SBS^{-1})$
  $= \det(S) \det(B) \det(S^{-1})$
  $= \det B.$ $\square$

- Similar matrices have the same eigenvalues. To show, we can show that both matrices have the same characteristic polynomial.

  Proof. We have

  $$\det(A - \lambda I) = \det(SBS^{-1} - \lambda SIS^{-1})$$
  $$= \det(S(B - \lambda I)S^{-1})$$
  $$= \det(S) \det(B - \lambda I) \det(S^{-1})$$
  $$= \det(B - \lambda I).$$

  $\square$

  This proves the result. However, it is not true that similar matrices have the same eigenvectors. For example, if $A = XDX^{-1}$, the eigenvectors of D are the standard basis vectors; while the eigenvectors of A are the columns of X.

Theorem. (Cayley-Hamilton). $\pi_A(A) = 0$. For example, $A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$, and $\pi_{(}A) = \lambda^2 - 2\lambda + 1$. Now, C-H states that $A^2 - 2A + I = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

We will now discuss matrix exponentials. Recall that we define

$$e^a = 1 + a + \frac{a^2}{2!} + \frac{a^3}{3!} + \dots .$$

In the case of a matrix, we define

$$e^{At} = I + At + \frac{A^2 t^2}{2!} + \frac{A^3 t^3}{3!} +$$

While there is no general algorithm to compute this, we can try diagonalizing. For instance, if D is diagonal, we can write

$$e^{Dt} = I + \operatorname{diag}(d_1 t, \dots, d_n t) + \dots + \operatorname{diag}(\frac{d_1^n}{n!}, \dots, \frac{d_n^n}{n!}) + \dots$$
$$= \operatorname{diag}(e^{d_1 t}, e^{d_2 t}, \dots, e^{d_n t}).$$

Claim. Now, if $A = XDX^{-1}$, then $e^{At} = Xe^{Dt}X^{-1}$.

Proof. We have

$$e^{At} = I + \sum_{k=1}^{\infty} \frac{A^k t^k}{k!} = I + \sum_{k=1}^{\infty} XD^k X^{-1} \frac{t^k}{k!}$$
$$= X \left[I + \sum_{k=1}^{\infty} \frac{D^k t^k}{k!} \right] X^{-1}.$$

This result follows. $\square$

Later in the class, we will note that matrix exponentials are useful to solve systems of linear ODEs.

If A is diagonalizable, we can prove the following results:

Claim. $\det(e^A) = e^A$. (HW)

Claim. $(e^A)^{-1} = e^{-A}$. (HW; follows immediately).

Remember A is not always diagonalizable. Recall that we proved that any symmetric matrix is diagonalizable. But we can show that any matrix is similar to a Jordan matrix.

Now, we will discuss probability vectors are Markov matrices.

Motivation. Often, you can have a system with many states. And with some probability, you have users that transition from state i to state j.

For instance, suppose we have smartphone users that either have iPhones, Samsung, Huawei. We can create a state transition matrix as follows:

$$M = \begin{pmatrix} 0.75 & 0.30 & 0.20 \\ 0.15 & 0.60 & 0.10 \\ 0.10 & 0.10 & 0.70 \end{pmatrix}.$$

Definition. A vector $v \in \mathbb{R}^n$ is a probability is $\sum_i v_i = 1$, with $v_i \geq 0$.

Definition. A matrix $M \in \mathbb{R}^{m \times n}$ is a Markov matrix if the columns of M are probability vectors.

Often, we are interested in the long-term behavior of the system.

Suppose we have an initial condition $v_0 = \begin{pmatrix} 1000 \\ 1000 \\ 1000 \end{pmatrix}$. In k years, the distribution will be $v_k = M^k v_0$.

The question is whether this sequence converges to some vector. That is, what is $\lim_{n \to \infty} M^n v_0$? This happens to be related to the eigenvalues of the matrix.

Let's now state a few properties of Markov matrices.

Claim. If M, N are Markov matrices, then MN is a Markov matrix. (Homework).

Claim. If M is a Markov matrix and v is a probability vector, then Mv is a probability vector. (Homework).

Theorem. For all λ_i eigenvalues of a Markov matrix M, we have $|\lambda_i| \geq 1$.

Claim. Any Markov matrix has eigenvalue 1. This is called the principal eigenvalue. (Homework). (Note: the corresponding eigenvector will be the limit.)

Now, suppose that a Markov matrix M is diagonalizable, so that $M = XDX^{-1}$. Now, $M^n = XD^nX^{-1}$. Then,

$$
\begin{aligned}
\lim_{n \to \infty} M^n &= \lim_{n \to \infty} XD^nX^{-1} \\
&= X \lim_{n \to \infty} D^nX^{-1} \\
&= X \operatorname{diag}(1, 0, \ldots, 0)X^{-1}. \qquad\qquad = v_1,
\end{aligned}
$$

where v_1 is the principal eigenvector.

12 Midterm review

12.1 Definitions

1. D2.1, Field.
2. D2.3, Vector space.
3. D2.6, subspace.
4. D2.10, linear dependence.
5. D2.12, span.
6. D2.14, basis.
7. D2.19, dimension.
8. D2.21, sum of subspaces.
9. D2.24, direct sum.
10. D3.1, linear maps.
11. D3.3 (image and kernel)
12. D3.9 (orthogonal complement)
13. D3.15 (injective and surjective linear maps)
14. 3.23 (bijective linear maps)
15. 3.16 (columnn and row ranks)
16. 4.1 (pseudoinverse) and its correctness (the bijectivity of the restriction, as proven in 3.17)

12.2 Theorems

1. E2.5 P1, 2. Show that $(\mathbb{R}^n, \mathbb{R})$ is a vector space. Show that $(\mathbb{R}^{m \times n}, \mathbb{R})$ is a vector space.

2. E2.8, P1, 2. Show that W (space of symmetric matrices) is a subspace of $\mathbb{R}^{m \times n}$.

   Let W be the space of orthogonal matrices. Then W is not a subspace of $\mathbb{R}^{m \times n}$ (just check scalar multiplication.

3. T2.22. Show that the sum of spaces is a subspace.

4. T2.26. Suppose $T = R \oplus S$. Then every $t \in T$ can be written uniquely in the form $t = r + s$, with $r \in R, s \in S$. Also, $\dim(T) = \dim R + \dim S$.

5. T3.4. Let $A : V \to W$ be a linear map. Then $\operatorname{Im} A, \operatorname{Ker} A$ are subspaces.

6. T 3.11, P2, 3. Let $R, S \subseteq \mathbb{R}^n$. Then:

   - $S \oplus S^\perp = \mathbb{R}^n$.
   - $(S^\perp)^\perp = S$.

7. T 3.12. Let $A : \mathbb{R}^n \to \mathbb{R}^n$. Then:

   - $\operatorname{Ker}(A)^\perp = \Im(A^T)$
   - $\operatorname{Im}(A)^\perp = \operatorname{Ker}(A^T)$.

8. T3.17. Let $A : \mathbb{R}^n \to \mathbb{R}^m$. Then $\dim \Im A = \dim \operatorname{Ker} A^\perp$. (Restricting a linear mapping to a bijection on fundamental subspaces; row and column ranks are equal).

9. T3.19, P1. Let $A, B \in \mathbb{R}^{n \times n}$. Then

$$0 \leq \operatorname{rank}(A + B) \leq \operatorname{rank} A + \operatorname{rank} B.$$

10. T3.20, P1, 2. Let $A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times p}$. Then

$$\operatorname{Im}(AB) \subseteq \operatorname{Im} A$$
$$\operatorname{Ker}(AB) \supseteq \operatorname{Ker}(A)$$

11. T3.21, P1, 3. Let $A \in \mathbb{R}^{m \times n}$. Then

$$\operatorname{Im}(A) = \operatorname{Im}(AA^T)$$
$$\operatorname{Ker}(A) = \operatorname{Ker}(A^T A).$$

12. T4.11. Let $A \in \mathbb{R}^{m \times n}$, and suppose $U \in \mathbb{R}^{m \times m}, V \in \mathbb{R}^{m \times n}$ are orthogonal. Then:

$$(UAV)^+ = V^T A^+ U^T.$$

12.3 Theoretical HW problems

1. (HW2.1) Determine whether the following are vector spaces over $\mathbb{R}$:

   - $\mathbb{R}$, w.r.t. $+, \cdot$.
   - $\mathbb{C}$, w.r.t. addition and multiplication.
   - , w.r.t. addition and multiplication.

2. (HW2.2)

   - Show that P_2 is a vector space over R.

- Let $1, t, 2t^2 - 1$ be the Chebyshev polynomials. Show they are a basis.
- Find components of $P_2(t)$ w.r.t. T_0, T_1, T_2.

3. (HW 2.3). Show that M'_n, M''_n are subspaces of $\mathbb{R}^{n \times n}$. Find dimension by finding a basis.

   Show that $\mathbb{R}^{n \times n}, i\mathbb{R}^{n \times n}$ are not subspaces of $\mathbb{C}^{n \times n}$.

4. (HW 2.4) Show that $L = \{p : p(-t) = p(t)\}, M = \{p : p(-t) = -p(t)\}$ are subspaces, and find their dimensions.

5. (HW 3.1) Show that $\mathbb{R}^{n \times n} = M'_n + \oplus M''_n$. Show that $P_2 = L \oplus M$.

6. (HW 3.4) Show that $M''_n = (M'_n)^\perp$, and that $M'_n = (M''_n)^\perp$.

7. (HW 3.6) Let R, S be subspaces of $\mathbb{R}^n$. Show that $R \subseteq S$ iff $S^\perp \subseteq R^\perp$.

8. (HW 4.1) Let $v_1, \ldots, v_n$ be orthonormal, and $A \in \mathbb{R}^{n \times n}$. Show that $Av_1, \ldots, Av_n$ are orthonormal iff the matrix A is orthogonal.

9. (HW 4.4) Let $A = \begin{pmatrix} 2 & 4 \\ 3 & 6 \end{pmatrix}$.

   - Calculate $B = \lim_{\delta \to 0} (A^T A + \delta I)^{-1} A^T$.
   - Find $\operatorname{Ker} A, \operatorname{Im} A, \operatorname{Ker} A^T, \operatorname{Im} A^T$.
   - Show that B is the pseudoinverse of A by erifying Definition 4.1.
   - Show that B is the pseudoinverse of A by verifying the condition of 4.2.

12.4 Computational HW problems

1. (HW3.2) For each mapping, state whether it is linear or not.

2. (HW3.3) Find the matrix of A with respect to the bases $U = (u_1, u_2, u_3)$ and $V = (v_1, v_2)$, where u_i, v_i are given.

3. (E3.2, P2) Define $L : \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$ so that $LX = MX$, for some fixed matrix M.

4. (E 3.2, P3 + HW3 P5). Let L be the differentiation operator w.r.t. the polnynomial p. Is L injective, surjective, bijective?

5. (HW 4.2) Find the four fundamental subspaces of the matrix $A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ (change numbers).

6. (HW 4.4) (See above).

12.5 Subtest A

Definitions.

1. Define a field.

2. Define a vector space.

3. Define a subspace.

4. Define the pseudoinverse (and it's correctness, by bijectivity of the restriction).

Theorems.

1. T2.26. Suppose $T = R \oplus S$. Then every $t \in T$ can be written uniquely in the form $t = r + s$, with $r \in R, s \in S$. Also, $\dim(T) = \dim R + \dim S$.

2. T3.4. Let $A : V \to W$ be a linear map. Then $\operatorname{Im} A, \operatorname{Ker} A$ are subspaces.

3. T 3.11, P2, 3. Let $R, S \subseteq \mathbb{R}^n$. Then:

   - $S \oplus S^\perp = \mathbb{R}^n$.

   - $(S^\perp)^\perp = S$.

4. T 3.12. Let $A : \mathbb{R}^n \to \mathbb{R}^n$. Then:

   - $\mathrm{Ker}(A)^\perp = \Im(A^T)$

   - $\mathrm{Im}(A)^\perp = \mathrm{Ker}(A^T)$.

5. T3.17. Let $A : \mathbb{R}^n \to \mathbb{R}^m$. Then $\dim \Im A = \dim \mathrm{Ker}\, A^\perp$. (Restricting a linear mapping to a bijection on fundamental subspaces; row and column ranks are equal).

6. T3.19, P1. Let $A, B \in \mathbb{R}^{n \times n}$. Then

$$0 \leq \mathrm{rank}(A + B) \leq \mathrm{rank}\, A + \mathrm{rank}\, B.$$

7. T4.11. Let $A \in \mathbb{R}^{m \times n}$, and suppose $U \in \mathbb{R}^{m \times m}, V \in \mathbb{R}^{m \times n}$ are orthogonal. Then:

$$(UAV)^+ = V^T A^+ U^T.$$

Theoretical HW.

1. (HW 2.3). Show that M_n', M_n'' are subspaces of $\mathbb{R}^{n \times n}$. Find dimension by finding a basis.

   Show that $\mathbb{R}^{n \times n}, i\mathbb{R}^{n \times n}$ are not subspaces of $\mathbb{C}^{n \times n}$.

2. (HW 3.1) Show that $\mathbb{R}^{n \times n} = M_n' + \oplus M_n''$. Show that $P_2 = L \oplus M$.

3. (HW 3.4) Show that $M_n'' = (M_n')^\perp$, and that $M_n' = (M_n'')^\perp$.

4. (HW 3.6) Let R, S be subspaces of $\mathbb{R}^n$. Show that $R \subseteq S$ iff $S^\perp \subseteq R^\perp$.

5. (HW 4.1) Let $v_1, \ldots, v_n$ be orthonormal, and $A \in \mathbb{R}^{n \times n}$. Show that $Av_1, \ldots, Av_n$ are orthonormal iff the matrix A is orthogonal.

6. (HW 4.4) Let $A = \begin{pmatrix} 2 & 4 \\ 3 & 6 \end{pmatrix}$.

   - Calculate $B = \lim_{\delta \to 0} (A^T A + \delta I)^{-1} A^T$.

   - Find $\mathrm{Ker}\, A, \mathrm{Im}\, A, \mathrm{Ker}\, A^T, \mathrm{Im}\, A^T$.

   - Show that B is the pseudoinverse of A by erifying Definition 4.1.

   - Show that B is the pseudoinverse of A by verifying the condition of 4.2.

Computational HW.

1. (HW3.3) Find the matrix of A with respect to the bases $U = (u_1, u_2, u_3)$ and $V = (v_1, v_2)$, where u_i, v_i are given.

2. (HW 4.2) Find the four fundamental subspaces of the matrix $A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ (change numbers).

12.6   Subtest A, Practice solve

Definitions.

1. Define a field.

   A field is a set with $(+, *)$, where:

   - $+$ is commutative, associative.

- $-x$ exists.

- 0 exists, with $x + 0 = x$.

Also,

- $*$ is commutative, associative.

- For nonzero x, x^{-1} exists.

- 1 exists, with $x \cdot 1 = x$.

Finally, $*$ distributes over addition.

2. Define a vector space.

   A vector space is a set with $(+, *)$, addition, scalar multiplication, where addition is defined over the set, and scalar multiplication is defined over the field.

   - $(V, +)$ is an abelian group.

   - V is closed under addition, and scalar multiplication.

   - There is a 0 element.

   - Additive inverses exist.

3. Define a subspace. A subset of a vector space that is closed under addition and scalar multiplication.

4. Define the pseudoinverse (and it's correctness, by bijectivity of the restriction).

   Defined as

$$\lim_{\delta \to 0} (A^T A + \delta I)^{-1} A^T.$$

   Alternatively, can verify properties from the book (verify).

Theorems.

1. T2.26. Suppose $T = R \oplus S$. Then every $t \in T$ can be written uniquely in the form $t = r + s$, with $r \in R, s \in S$. Also, $\dim(T) = \dim R + \dim S$.

   $T = R \oplus S$ means $R + S = T$ and $R \cap S = \{0\}$. Suppose some $t = r_1 + s_1 = r_2 + s_2$. Then, $r - 1 - r_2 = s_1 - s_2$, but this contradicts that fact that $R \cap S = \{0\}$, since this forces $r_1 = r_2; s_1 = s_2$.

2. T3.4. Let $A : V \to W$ be a linear map. Then $\operatorname{Im} A, \operatorname{Ker} A$ are subspaces. Easy.

3. T 3.11, P2, 3. Let $R, S \subseteq \mathbb{R}^n$. Then:

   - $S \oplus S^\perp = \mathbb{R}^n$.

   - $(S^\perp)^\perp = S$.

   First property - show that intersection is just 0, which is easy. Then, show that they sum to $\mathbb{R}^n$. Can do this with

   Note that $S^\perp = \{t \in \mathbb{R}^n | \langle s, t \rangle = 0, s \in S\}$. Not super sure how to conclude.

4. T 3.12. Let $A : \mathbb{R}^n \to \mathbb{R}^n$. Then:

   - $\operatorname{Ker}(A)^\perp = \operatorname{Im}(A^T)$

   - $\operatorname{Im}(A)^\perp = \operatorname{Ker}(A^T)$.

   If $w \cdot v = 0$, and $Av = 0$, then $A^T v = w$. Not super sure how to conclude.

5. T3.17. Let $A : \mathbb{R}^n \to \mathbb{R}^m$. Then $\dim \operatorname{Im} A = \dim \operatorname{Ker} A^\perp$. (Restricting a linear mapping to a bijection on fundamental subspaces; row and column ranks are equal).

   Corollary of the previous part.

6. T3.19, P1. Let $A, B \in \mathbb{R}^{n \times n}$. Then

$$0 \le \operatorname{rank}(A + B) \le \operatorname{rank} A + \operatorname{rank} B.$$

   Not sure.

7. T4.11. Let $A \in \mathbb{R}^{m \times n}$, and suppose $U \in \mathbb{R}^{m \times m}, V \in \mathbb{R}^{m \times n}$ are orthogonal. Then:

$$(UAV)^+ = V^T A^+ U^T.$$

   Not sure.

Theoretical HW.

1. (HW 2.3). Show that M_n', M_n'' are subspaces of $\mathbb{R}^{n \times n}$. Find dimension by finding a basis.

   Not sure, look at solutions.

   Show that $\mathbb{R}^{n \times n}, i\mathbb{R}^{n \times n}$ are not subspaces of $\mathbb{C}^{n \times n}$.

   This part is easy.

2. (HW 3.1) Show that $\mathbb{R}^{n \times n} = M_n' \oplus M_n''$. Show that $P_2 = L \oplus M$.

   To argue first part, just need to show that intersection is 0, and that the sum sums to all possible matrices.

3. (HW 3.4) Show that $M_n'' = (M_n')^\perp$, and that $M_n' = (M_n'')^\perp$.

   Requires analyzing the term structure.

4. (HW 3.6) Let R, S be subspaces of $\mathbb{R}^n$. Show that $R \subseteq S$ iff $S^\perp \subseteq R^\perp$.

   Not hard, but don't remember the full argument.

5. (HW 4.1) Let $v_1, \ldots, v_n$ be orthonormal, and $A \in \mathbb{R}^{n \times n}$. Show that $Av_1, \ldots, Av_n$ are orthonormal iff the matrix A is orthogonal.

6. (HW 4.4) Let $A = \begin{pmatrix} 2 & 4 \\ 3 & 6 \end{pmatrix}$.

   - Calculate $B = \lim_{\delta \to 0}(A^T A + \delta I)^{-1} A^T$.

     Easy calculation.

   - Find $\operatorname{Ker} A, \operatorname{Im} A, \operatorname{Ker} A^T, \operatorname{Im} A^T$.

     Easy.

   - Show that B is the pseudoinverse of A by verifying Definition 4.1.

     Easy, but annoying (need to recall definition).

   - Show that B is the pseudoinverse of A by verifying the condition of 4.2.

     Easy but annoying (need to recall the criteria).

Computational HW.

1. (HW3.3) Find the matrix of A with respect to the bases $U = (u_1, u_2, u_3)$ and $V = (v_1, v_2)$, where u_i, v_i are given.

   Easy, but annoying. Need to remember matrix product form.

2. (HW 4.2) Find the four fundamental subspaces of the matrix $A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ (change numbers).

Straightforward. Need to calculate $\operatorname{Im} A, \operatorname{Ker} A, \operatorname{Im} A^T, \operatorname{Ker} A^T$.

Review: subtest A.

Definitions.

1. Define a vector space.

   A field F with two operations $+ : V \times V \to V, * : F \times V$ that satisfy:

   - $(V, +)$ is an abelian group.

   - Associativity (mult) in field times vector.

   - Distributivity (both dirs).

   - $1 \cdot v = v$.

2. Define the pseudoinverse. Defined as

$$\lim_{\delta \to 0} (A^T A + \delta^2 I)^{-1} A^T.$$

   $B : \operatorname{Ker}(A)^\perp \to \operatorname{Im} A$ is bijective. Need to prove this.

Theorems.

1. T 3.11, P2, 3. Let $R, S \subseteq \mathbb{R}^n$. Then:

   - $S \oplus S^\perp = \mathbb{R}^n$.

   - $(S^\perp)^\perp = S$.

   Easiest way to do this is to think about the geometry in $\mathbb{R}^n$. Let v_i be orthonormal basis for S. Then we can write

$$x_1 = \sum_{i=1}^{k} (x^T v_i) v_i$$
$$x_2 = x - x_1.$$

   Note that $x_2^T v_j = x^T v_j - x_1^T v_j = x^T v_j - x^T v_j = 0$, so x_2, v_j are orthogonal for any j, and thus to any linear combination of these vectors. Thus, x_2 is orthogonal to S.

   Note that $\dim S = \dim(S^\perp)^\perp$. Now, $W \subset W^{\perp\perp}$. QED.

2. T 3.12. Let $A : \mathbb{R}^n \to \mathbb{R}^n$. Then:

   - $\operatorname{Ker}(A)^\perp = \operatorname{Im}(A^T)$

   - $\operatorname{Im}(A)^\perp = \operatorname{Ker}(A^T)$.

   Take $v \in \operatorname{Ker} A$. Then $Av = 0$. Implies that $y^T Av$ for all y. But this means $(A^T y)^T x = 0$ for all y, so we have orthogonality.

3. T3.17. Let $A : \mathbb{R}^n \to \mathbb{R}^m$. Then $\dim \operatorname{Im} A = \dim \operatorname{Ker} A^\perp$. (Restricting a linear mapping to a bijection on fundamental subspaces; row and column ranks are equal).

   Let $T : \operatorname{Ker} A^\perp \to \operatorname{Im} A$. Then T is bijective (not too hard to see.)

4. T3.19, P1. Let $A, B \in \mathbb{R}^{n \times n}$. Then

$$0 \leq \operatorname{rank}(A + B) \leq \operatorname{rank} A + \operatorname{rank} B.$$

Just concat the bases, and argue that every row of $(A + B)$ works as a linear combination of the respective bases.

5. T4.11. Let $A \in \mathbb{R}^{m \times n}$, and suppose $U \in \mathbb{R}^{m \times m}, V \in \mathbb{R}^{m \times n}$ are orthogonal. Then:

$$(UAV)^+ = V^T A^+ U^T.$$

Just verify that it satisfies the four Penrose conditions, namely $ABA = A, BAB = B, (AB)^T = AB, (BA)^T = BA$.

Theoretical HW.

1. (HW 2.3). Show that M_n', M_n'' are subspaces of $\mathbb{R}^{n \times n}$. Find dimension by finding a basis.

   $n(n+1)/2$, just argue by decomposition.

2. (HW 3.4) Just analyze the term structure. Not hard.

3. (HW 3.6) Let $s' \in S^\perp$. We must have $\langle s, s' \rangle = 0$ for all $s \in S$. Then $\langle s, s' \rangle = 0$ for all $s \in R$ in particular. Hence $s' \in S^\perp$ implies $s' \in R^\perp$. QED.

4. Suppose $v_1, \ldots, v_n$ is orthonormal. Show that $Av_1, \ldots, Av_n$ are orthonormal if the matrix A is orthogonal.

   Consider $\langle Av_i, Av_j \rangle = v_i^T A^T A v_j$. Argue each way.

Computational HW.

1. Just do $B = V^{-1} AU$, where U is first basis, and V is second basis.

12.7   Subtest B

Thms.

1. T3.17. Let $A : \mathbb{R}^n \to \mathbb{R}^m$. Then $\dim \operatorname{Im} A = \dim \operatorname{Ker} A^\perp$. (Restricting a linear mapping to a bijection on fundamental subspaces; row and column ranks are equal).

HWT.

1. Dimension of M_n'.

2. $v_1, \ldots, v_n$ orthonormal, and $A \in \mathbb{R}^{n \times n}$. $Av_1, \ldots, Av_n$ ortho if A is ortho.

HWC.

- Find matrix of A w.r.t. bases U and V. $F = V^{-1} AU$.

13   Lecture 16: 7-22-19

Recall that if $v, w \in \mathbb{R}^n$, we can define the standard inner product as

$$\langle v, w \rangle = v^T w.$$

Definition. In general, an inner product is a map $V \times V \to \mathbb{R}$. It must satisfy:

- Positive definite. So that $\langle v, v \rangle \geq 0$, and $\langle v, v \rangle = 0$ iff $v = 0$.

- Symmetry. We must have $\langle v, w \rangle = \langle w, v \rangle$ for all $v, w \in V$.

- Linearity. We must have $\langle v, \alpha w_1 + \beta w_2 \rangle = \alpha \langle v, w_1 \rangle + \beta \langle v, w_2 \rangle$ for all $v, w_1, w_2 \in V, \alpha, \beta \in \mathbb{R}$.

Example. Consider the vector space with P^n, where P^n denotes the space of degree n polynomials with real coefficients. We can define an inner product as follows:

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

Here, we note that

$$\langle f, f \rangle \geq 0,$$

and it satisfies all the axioms:

- $\int_{-1}^{1} f^2(x)\, dx = 0$.
- $\int_{-1}^{1} f(x)g(x) = \int_{-1}^{1} g(x)f(x)\, dx$.
- It satisfies linearity, since,

$$\langle f, \alpha g_1 + \beta g_2 \rangle = \int_{-1}^{1} f(x)\left[\alpha g_1(x) + \beta g_2(x)\right]$$

$$= \alpha \int_{-1}^{1} f(x)g_1(x)\, dx + \beta \int_{-1}^{1} f(x)g_2(x)\, dx$$

$$= \alpha \langle f, g_1 \rangle + \beta \langle f, g_2 \rangle.$$

Definition. We say that a map $V \times V \to \mathbb{C}$ is a complex inner product if the following axioms hold:

- *Positive definite.* That is, $\langle v, v \rangle \geq 0$ for all $v \in V$, and $\langle v, v \rangle = 0$ iff $v = 0$.
- *Hermitian symmetry.* For all v, w, we have $\langle v, w \rangle = \overline{\langle w, v \rangle}$.
- *Hermitian linearity.* We have

$$\langle v, \alpha w_1 + \beta w_2 \rangle = \overline{\alpha}\langle v_1, w \rangle + \overline{\beta}\langle v_2, w \rangle.$$

For example, note that $\langle v, w \rangle = v^H w$ is a complex inner product.

We now discuss vector norms. Given a vector space V over a field $\mathbb{F}$, a norm is a map $|| \cdot || : V \to \mathbb{R}$ that satisfies the following properties:

- *Positive definite.* We must have $||v|| \geq 0$ for all v, and $||v|| = 0$ iff $v = 0$.
- *Linearity.* This means that $||\alpha v|| = |\alpha|||v||$ for all $v \in V, \alpha \in \mathbb{R}$.
- *Triangle inequality.* We have that $||v + w|| \leq ||v|| + ||w||$.

Here are some examples of norms. If $v \in \mathbb{R}^n$, then

- $||v||_2 = \left(\sum_{i=1}^{n} v_i^2\right)^{\frac{1}{2}}$.
- $||v||_p = \left(\sum_{i=1}^{n} v_i^p\right)^{\frac{1}{p}}$.
- $||v||_1 = \left(\sum_{i=1}^{n} |v_i|\right)$ (also known as the Manhattan norm).
- $||v||_\infty = \max_i |v_i|$.

There are some important inequalities we can state related to these norms (the proofs can be found in the book).

Theorem. (Hölder's Inequality). If $\frac{1}{p} + \frac{1}{q} = 1$, then

$$|v^H w| \leq ||v||_p ||w||_q.$$

A special case of Hölder's Inequality is the Cauchy-Schwarz inequality. That is, when $p = q = 2$.

Theorem. (Cauchy-Schwarz inequality).

$$|v^H w| \le ||v||_2 ||w||_2.$$

Now, given a vector inner product, we can always define a norm induced by the inner product. Suppose $\langle \cdot, \cdot \rangle$ is an inner product from $V \times V \to \mathbb{R}$. Then we can always define a norm induced by this inner product by

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

But, it is not true that all inner products come from a norm.

Definition. Suppose that $|\cdot|_\star$ and $||\cdot||_*$ are norms of V. We say that they are equivalent if there exists constants $c_1, c_2 > 0$ such that for any $v \in V$ we have

$$c_1 ||v||_\star \le ||v||_* \le c_2 ||v||_\star.$$

For motivation, we note that if norms are equivalent, then they will preserve topological properties in the space. For example, if a function converges in the 1 norm, it will converge in the 2 norm.

Example. We have for n dimensional vectors $v \in \mathbb{R}^n$, we have

$$||v||_2 \le ||v||_1 \le \sqrt{n} ||v||_2,$$

so the $||v||_1$ and $||v||_2$ are equivalent.

We note that the right inequality is a consequence of Cauchy-Schwarz.

Example. We have for n dimensional vectors $v \in \mathbb{R}^n$, we have

$$||v||_\infty \le ||v||_1 \le n ||v||_\infty.$$

Proof. We have

$$||v||_1 = \sum_{i=1}^{n} |v_i| \le \sum_{i=1}^{n} \max |v_i| = n ||v||_\infty.$$

This constant is sharp because equality is achieved when $v = (1, 1, \ldots, 1)$. The opposite inequality is obvious. Clearly, $\max |v_i| \le \sum_{i=1}^{n} |v_i|$. This inequality is sharp by taking a vector $(1, 0, \ldots, 0)$. □

Now, we continue to discuss matrix norms.

Suppose $A \in \mathbb{R}^{m \times n}$. Now, while thinking about the map $v \mapsto Av$, we can define a special type of matrix norm on it.

Definition. Suppose we have $||\cdot||_*$ on $\mathbb{R}^n$; and $||\cdot||_\star$ on $\mathbb{R}^m$. Then, we can define an induced matrix norm as follows:

$$||A||_{*,\star} = \max_{v \neq 0} \frac{||Av||_\star}{||v||_*}.$$

This is also known as the operator norm.

Example. Suppose $A = I_n$, and we consider the 2-norms on both the domain and the co-domain.

$$||I||_{2,2} = \max_{v \neq 0} \frac{||v||_2}{||v||_2} = 1.$$

Example. Suppose now $A = I_n$, and we use the infinity norm in the domain and the 1-norm in the co-domain. That is,

$$||I||_{\infty,1} = \max_{v \neq 0} \frac{||v||_1}{||v||_\infty} = n.$$

This comes from the inequality $||v||_1 \leq n||v||_\infty$ we proved earlier.

14 Lecture 17: 7-23-19

Definition. Let V be a vector space with $V = X \oplus Y$. Define $P_{X,Y} : V \to X \subseteq V$ by

$$P_{X,Y}v = x,$$

for all $v \in V$. Then $P_{X,Y}$ is called the (oblique) projection on X along Y.

Definition. We say that P is an orthogonal projection if $Y = X^\perp$.

Theorem. A linear transformation P is a projection iff $P^2 = P$.

Theorem. P is a projection iff $I - P$ is also a projection.

The reason we care about orthogonal projections is that the matrices associated with them are orthogonal.

Theorem. $P \in \mathbb{R}^{n \times n}$ is the matrix for an orthogonal projection onto $\text{Im}(P)$ iff $P^2 = P = P^T$. (In fact, P is a little more than orthogonal, since we require $P^2 = P$).

We will skip the proof. The first part, that $P^2 = P$ follows from the previous result. The trickiest part is the transpose condition (this is proven in the book, see 7.5).

We will now discuss properties of projections.

Theorem. Let P be a matrix for orthogonal projection. Then

$$||Pv|| \leq ||v||,$$

where $|| \cdot ||$ is any vector name on V.

Proof. This follows from the Pythagorean theorem. By this theorem, if $v \perp w$, then $||v + w||^2 = ||v||^2 + ||w||^2$ (this applies to any norm). We know that $v = Pv + (I - P)v$. Now, $Pv \perp (I - P)v$, and so by this theorem

$$||v||^2 = ||Pv||^2 + ||(I - P)v||^2,$$

implying $||Pv||^2 \leq ||v||^2$. $\qquad\square$

Example. Project onto the span of v in $\mathbb{R}^2$, so that $P_v : \mathbb{R}^2 \to \mathbb{R}^2$. We can write

$$P_v(u) = \frac{u \cdot v}{v \cdot v} + v.$$

Suppose tha $tv_1, \ldots, v_n$ are orthogonal, and $X = \text{span}\{x_1, \ldots, x_n\}$. Then

$$\text{Proj}_X(u) = \sum_{i=1}^n \text{Proj}_{v_i}(u) = \sum_{i=1}^n \frac{u \cdot v_i}{v_i \cdot v_i} v_i.$$

Suppose that $v_1, \ldots, v_n$ in $\mathbb{R}^m$, $m \geq n$ are orthonormal. Suppose that $e_1, \ldots, e_m$ is the standard basis. Suppose that $P : \mathbb{R}^m \to \mathbb{R}^m$ is the linear map of orthogonal projections on $\text{span}\{v_1, \ldots, v_n\}$. How do we find the projection matrix? We just put the projection of the i-th basis vector in the i-th column of the matrix.

Now, we discuss the Gram-Schmidt orthogonalization process.

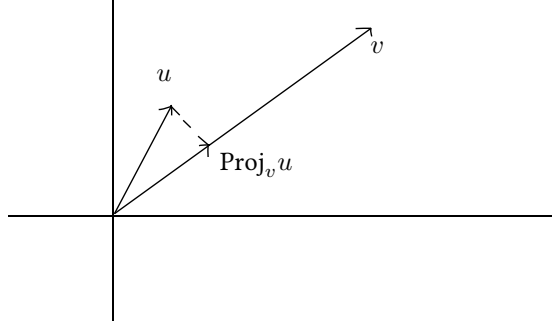Suppose that $w_1, \ldots, w_n$ are linearly independent. Then we apply this process by following these steps:

Figure 1: Projection of u onto v

1. Normalize w_1, by writing

$$v_1 = \frac{w_1}{||w_1||}.$$

2. w_2 is not in general orthogonal to w_1. So, we set

$$q_2 = w_2 - \text{Proj}_{v_1} w_2,$$

and here q_2 is orthogonal to w_1.

Now, normalize, to obtain $v_2 = \frac{q_2}{||q_2||}$.

3. Keep going, so

$$q_3 = w_3 - \text{Proj}_{v_1} w_3 - \text{Proj}_{v_2} q_3.$$

Now,

$$v_3 = \frac{q_3}{||q_3||}.$$

4. In general, if we continue this process, we obtain

$$q_{k+1} = w_{k+1} - \sum_{i=1}^{k} \text{Proj}_{v_i} w_{k+1}$$

$$v_{k+1} = \frac{q_{k+1}}{||q_{k+1}||}.$$

And then, the resulting set of v_i will have the same span of the w_i. To see this, note that the w_i are linearly independent and form a basis; similarly since the v_i are orthogonal, they must span the space for dimensionality reasons.

15   Lecture 18: 7-24-19

We define the QR factorization as follows. Given a matrix $A \in \mathbb{R}^{m \times n}$, where $\text{rank}(A) = n$, we can write $A = QR$, where $Q \in \mathbb{R}^{m \times m}$, R is an upper triangular $m \times n$ matrix.

To compute the QR decomposition, we can use Gram Schmidt to compute $q_1, \ldots, q_n \in \mathbb{R}^m$ which are orthonormal. And we can compute scalars $r_{ij} = \langle v_j, q_i \rangle$. Suppose we define $Q^1 = \begin{pmatrix} q_1 & q_2 & \ldots q_n \end{pmatrix}$, $R^1_{ij} = (r_{ij})$. Then we have that $A = Q^1 R^1$.

Now, suppose that $q_{n+1}, \ldots, q_m$ are orthonormal such that $q_1, \ldots, q_m$ form an orthonormal basis for $\mathbb{R}^m$. Then we can write $Q = \begin{pmatrix} q_1 & q_2 & \ldots & q_m \end{pmatrix}$. So we can write

$$A = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \begin{pmatrix} R_1 \\ 0 \end{pmatrix} = Q_1 R_1.$$

One application of QR decomposition is that they allow you to solve systems of equations. Suppose we have $Ax = b$, this is equivalent to $QRx = b$, so $Rx = Q^{-1}b$.

To compute $q_{n+1}, \ldots, q_m$ - we can look at the Span of $\{q_1, \ldots, q_n\}$ which is not equal to $\mathbb{R}^m$. To pick the rest of the vectors: pick v_{n+1} such that $v_{n+1} \in \mathbb{R}^m$, but v_{n+1} is not in the span. Then: just re-apply Gram-Schmidt to compute the rest of the q_i.

We will now discuss least-squares problems, which are used when we cannot solve a system explicitly. For instance, we might have a system $Ax = b$ that cannot be solved explicitly. The most standard case is linear regression - when we want to minimize the L_2 error of the predictions vs. the data. For example, if we are trying to find a function that takes heights to weights, we might have the dataset,

$$\begin{pmatrix} h = 175 & w = 50 \\ 180 & 70 \\ 164 & \\ 55 & \end{pmatrix}.$$

and we want to take $h(w) = aw + b$, and learn constants a, b.

Pictorially, if the blue dots are the data points, we want to find a line that minimizes the sum of the d_i^2.

Definition. The solution of the least squares problem is x_0 such that $||Ax_0 - v||^2$ is minimized.

We can use projections to relate Ax_0 are v. We note that $Ax_0 \in \text{Im}(A)$. We want to pick v where v is the orthogonal projection onto $\text{Im } A$. Namely, $Ax_0 = \text{Proj}_{\text{Im } A} v$.

The key point is that the vector $v - Ax_0$ is orthogonal to $\text{Im } A$. This means that $(v - Ax_0) \in \text{Ker } A^T$, so $A^T(v - Ax_0) = 0$. Thus, $A^T V = A^T Ax_0$.

Now, suppose that A has linearly independent columns. Then $A^T A$ is invertible. Givne the equation $A^T v = A^T Ax_0$, we can find $x_0 = (A^T A)^{-1} A^T v$.

Example (linear regression). Suppose we have a set of points (x_i, y_i) for $1 \le i \le N$, and we want to solve $y = \alpha x + \beta$. And we can build a matrix $A = \begin{pmatrix} x_1 & 1 \\ \ldots & \\ x_n & 1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} y_1 \\ \ldots \\ y_n \end{pmatrix}$.

Example (polynomial regression). Suppose we have $y = \alpha_k x^k + \alpha_{k-1} x^{k-1} + \cdots + \alpha_0$. In this case, the system of equations looks like

$$y_i = \alpha_k x^k + \alpha_{k-1} x^{k-1} + \cdots + \alpha_0.$$

In this case, the unknowns are the α_i's. Here, we can build a bigger matrix, where

$$\begin{pmatrix} x_1^k & x_1^{k-1} & \ldots & 1 \\ x_2^k & x_2^{k-1} & \ldots & 1 \\ \ldots & & & \\ x_n^k & x_n^{k-1} & \ldots & 1 \end{pmatrix} \begin{pmatrix} \alpha_k \\ \alpha_{k-1} \\ \ldots \\ \alpha_0 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix}.$$
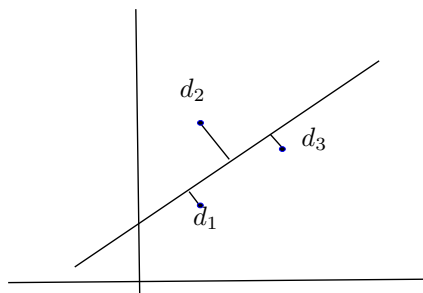
Figure 2: Linear regression

Now, this is a system $Ax = v$, you want to find the solution that minimizes the squared error - so this is equivalent to the setting from before.

16   Lecture 21: 7-26-19

End of last week, we were doing last least squares problems, and yesterday we discussed SVDs.

Recall that the SVD is like a generalization of the eigenvector / eigenvalue decomposition (since not every matrix has an eigenvector / eigenvalue decomposition).

Remember that we can write $A = UDV^T$, where U, V are orthogonal, and D is diagonal (with added zeros as needed).

To obtain the SVD, we perform the following algorithm:

- Compute eigenvalues for $A^T A$. We are guaranteed to find n eigenvalues; exactly r are nonzero.

- Compute eigenvectors $v_1, \ldots, v_n$ (require orthonormal). The spectral theorem allows us to obtain a full basis of eigenvectors, since $A^T A$ is a symmetric matrix.

- Look at the vectors $Av_1, \ldots, Av_n$. Then $||Av_i|| = \lambda_i$, so exactly r of Av_i are not zero. We proved yesterday that the Av_i are orthogonal.

This allows us to compute the SVD matrices. We obtain:

$$
\begin{aligned}
V &= \begin{pmatrix} v_1 & v_2 & \ldots & v_n \end{pmatrix} \\
D &= \operatorname{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_r}, 0, \ldots, 0) \\
U &= \begin{pmatrix} \frac{Av_1}{\sqrt{\lambda_1}} & \frac{Av_2}{\sqrt{\lambda_2}} & \ldots & \frac{Av_r}{\sqrt{\lambda_r}} & u_{r+1} & \ldots & u_n \end{pmatrix}.
\end{aligned}
$$

Now, the $u_{r+1}, \ldots, u_n$ are redundant (because the the right portion of the matrix gets zeroed out be the zero entries in D).

If we write

$$U = \begin{pmatrix} U_1 & U_2 \end{pmatrix}, D = \begin{pmatrix} D_1 & 0 \\ 0 & 0 \end{pmatrix}, V = \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix},$$

we can compute a reduced SVD as $A = U_1 D_1 V_1^T$.

Importantly, here D is unique; but U and V are not unique. However, the eigenspaces associated with U and V are unique.

Geometrically, this is really cool, because it means that you can express any matrix A as the product of an isometry (V^T), a scaling with D, and another isometry (U). These isometries can be reflections, rotations, or rotoreflection.

Properties of vector norms. Recall that $|| \cdot || : V \to \mathbb{R}$ is a function that satisfies the following properties:

- *Positive definite.* $\langle v, v \rangle \geq 0$ for all v, with equality iff $v = 0$.

- *Linearity.* $||av|| = |a|||v||$ for all $v \in V, a \in \mathbb{R}$.

- *Triangle inequality.* We have that $||v + w|| \leq ||v|| + ||w||$.

Definition. We say that a matrix norm is a norm on the vector space $\mathbb{R}^{m \times n}$.

Let's consider a few examples.

- Frobenius norm, defined as

$$||A||_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \right)^{\frac{1}{2}} = (\sum_{k=1}^{r} \sigma_k^2)^{1/2}.$$

- Induced p-norm. We can define:

$$||A||_p = \max_{x \neq 0} \frac{||Ax||_p}{||x||_p}.$$

- Induced pq-norm. We can define

$$||A||_{p,q} = \max_{x \neq 0} \frac{||Ax||_q}{||Ax||_p}.$$

For homework, we will show that

$$||A||_2 = \sigma_1(A),$$

i.e. the largest singular value.

Now, we can prove that different norms are equivalent with matrix norms as well.

Definition. We say that $|| \cdot ||_\alpha, || \cdot ||_\beta$ on $\mathbb{R}^{m \times n}$ are equivalent if there exists $c_1, c_2 > 0$ such that for all matrices $A \in \mathbb{R}^{m \times n}$, we have

$$c_1 ||A||_\alpha \leq ||A||_\beta \leq c_2 ||A||_\alpha.$$

To prove matrix norms are equivalent, we can make heavy use of the equivalence of vector norms.

17   Lecture 22: 7-31-19

Note that
$$||\lambda \cdot x||_\star = |\lambda|||x||_\star.$$

Now, we note that we can compute matrix norms over the unit ball instead of the whole vector space, since:
$$\max_{x \neq 0} \frac{||Ax||_\star}{||x||_\star} = \max_{||y||_\star = 1} ||Ay||_\star.$$

Now, if $||A||_1 = \max_{||x|| \neq 0} \frac{||Ax||_1}{||x||_1} = \max_{||x||=1} ||Ax||$. This latter problem is maximizing a continuous function on a compact set - so there must be a maximum.

Claim. We have
$$||A||_1 = \sum_{1 \leq j \leq m} \sum_{i=1}^{n} |a_{ij}|; \qquad \text{max col sum.}$$

Proof. Suppose $A_1, \ldots, A_m$ are the column of A. We want to look at
$$||Ax||_1 = \max_{||x||=1} \sum_{j=1}^{m} x_j A_j$$
$$\leq \sum_{j=1}^{m} |x_j A_j|_1$$
$$= \sum_{j=1}^{m} |x_j|||A_j||_1$$
$$\leq \sum_{j=1}^{m} |x_j| \left(\max_{1 \leq k \leq m} ||A_k|_1 \right)$$
$$= \left(\max_{1 \leq k \leq m} ||A_k||_1 \right) ||x||_1.$$

This shows that for all x, we have:
$$\frac{||Ax||_1}{||x||_1} \leq \max_{1 \leq k \leq n} ||A_k||_1.$$

Since this is true for all x, it must be true for the maximum x, that is:
$$||A||_1 \leq \max_{1 \leq k \leq m} ||A_k||_1.$$

Now, note that for any x_0 fixed, we have
$$\frac{||Ax_0||_1}{||x_1||_1} \leq ||A||_1.$$

Let l be such that $||A_l||_1 \leq ||A_k||$ for all $1 \leq k \leq m$.

Now, take $x_0 = e_l$, the l-ith basis vector. Then:
$$\frac{||Ax_0||_1}{||x_0||_1} = \frac{||Al||_1}{||l||_1} = \max_{1 \leq k \leq n} ||A_k||_1 = ||A_l||_1.$$

$\square$

Claim. We have

$$||A||_\infty = \text{maximum row sum}$$
$$||A||_2 = \sigma_1(A); \qquad \text{maximum singular value}$$
$$||A||_{pq} = \left(\sum_{j=1}^{n} \left(\sum_{i=1}^{m} |a_{ij}|^p \right)^{q/p} \right)^{1/q}.$$

Definition. We say that a matrix norm is consistent if

$$||AB|| \le ||A||||B||.$$

Claim. All induced norms are consistent.

Proof. To show this, we need to prove a bunch of triangle inequalities. We have:

$$\frac{||Ax||}{||x||} \le \max_{y \ne 0} \frac{||Ay||}{||y||} = ||A||.$$

Now, this implies that $||Ax|| \le ||A|| \cdot ||x||$. To continue, we note that

$$||ABx|| \le ||A||||Bx|| \le ||A||||B||||x||.$$

Now, dividing both sides by the norm of x, we obtain for $x \ne 0$,

$$\frac{||ABx||}{||x||} \le ||A||||B||,$$

thus,

$$\max_{x \ne 0} \frac{||ABx||}{||x||} = ||AB|| \le ||A||||B||.$$

$\square$

In the HW - problem asks whether Frobenius norm is induced by a vector norm. Hint: show that Frobenius norm is not consistent, therefore it is not induced.

Note that all p-norms are equivalent. This is easy to show, if we just note that

$$\left(\sum |x_i|^p \right)^{1/p} \le \left(\sum \max |x_i|^p \right)^{1/p} = (n||x||_\infty^p)^{1/p}$$

Also,

$$||x||_\infty^p = \max |x_i|^p \le \sum_{i=1}^{n} |x_i|^p = ||x||_p^p.$$

The result follows.

Tomorrow, we will discuss the equivalence of matrix norms.

18 Lecture 23: 8-1-19

Yesterday, we showed that

$$||A||_1 = \max_{x \neq 0} \frac{||Ax||_1}{||x||_1} = \text{max col sum.}$$

Today, we note that

$$||A||_\infty = \max_{x \neq 0} \frac{||Ax||_\infty}{||x||_\infty} = \text{max row sum.}$$

Proof. If A_i is the i-th row of A, we have

$$||Ax||_\infty = \max_{1 \leq i \leq n} \langle A_i, x \rangle$$

$$= \max_{1 \leq i \leq n} \left| \sum_{j=1}^{m} A_{ij} x_j \right|$$

$$\leq \max_{1 \leq i \leq n} \sum_{j=1}^{m} |A_{ij}||x_j|$$

$$\leq \max_{1 \leq i \leq n} \sum_{j=1}^{m} |A_{ij}|||x||_\infty.$$

Then,

$$||A||_\infty \leq \max_{1 \leq i \leq n} \sum_{j=1}^{n} |A_{ij}|.$$

And this is achieved if you take x to be $\text{sgn}(A_m)$, where A_m is the maximum row.

(Note: this is part of the homework). □

Now, we will show that

$$||A||_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |A_{ij}|^2 \right)^{1/2} = \left[\sum_{i=1}^{r} \sigma_i(A) \right]^{1/2}.$$

Proof. Note that if P is orthogonal, we have $||PA||_F = ||A||_F$. This follows since orthogonal matrices preserve 2-norms.

Now, suppose $A = VDQ^H$ is an SVD. By SVD, we know that V and Q^H are orthogonal. So, $||A||_F = ||D||_F = \left[\sum_{i=1}^{r} \sigma_i(A)^2 \right]^{1/2}$. □

We will prove that the infinity norm is equivalent to the 1 norm.

Proof. Using the results for vectors that

- $||x||_\infty \leq ||x||_1$
- $||x||_1 \leq n||x||_\infty.$

Now, for a fixed vector x, we have that:

$$\frac{||Ax||_\infty}{||x||_\infty} \leq \frac{||Ax||_1}{\frac{1}{n}||x||_1}.$$

Now, we have that

$$\frac{||Ax||_\infty}{||x||_\infty} \leq \frac{n||Ax||_1}{||x||_1} \leq n \cdot \max_{y \neq 0} \frac{||Ay||_1}{||y||_1} = n \cdot ||A||_1.$$

Now, we have equality when $A = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & & & \end{pmatrix}$, i.e. the matrix with the first row all 1s, and the rest of the matrix 0.

Now, we can similarly prove that $||A||_1 \leq m||A||_\infty$, where equality is achieved when $A = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & \\ \vdots & & & \\ 1 & 0 & \cdots & 0 \end{pmatrix}$, that is, the first column is all 1s, and the rest is 0s.

In the homework, we need to show that 2 norms of matrices are equivalent to infinity norms of matrices. $\quad\square$

19 Lecture 24: 8-6-19

Definition. We define a Jordan block of size n for eigenvalue λ as

$$J_n(\lambda) = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}.$$

Theorem. For any complex valued matrix A, there exist X where $A = XJX^{-1}$, where J is a Jordan matrix.

We will discuss an algorithm to compute the JCF.

Example. Suppose $A = \begin{pmatrix} 4 & 1 \\ -1 & 6 \end{pmatrix}$. To diagonalize this, we perform the following steps.

First, we compute the eigenvalues: we get

$$\det(A - \lambda I) = \lambda^2 - 10\lambda + 25 = (\lambda - 5)^2 = 0,$$

so $\lambda = 5$ with multiplicity 2.

Next, we find eigenvectors by solving $Av = \lambda v$, that is $v \in \text{Ker}(A - \lambda I)$. Solving, we get the eigenvector $v = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

If we look at the product

$$A \begin{pmatrix} v_1 & v_2 \end{pmatrix} = \begin{pmatrix} v_1 & v_2 \end{pmatrix} = \begin{pmatrix} 5 & 1 \\ 0 & 5 \end{pmatrix}.$$

Equating columns, we obtain

$$Av_1 = 5v_1$$
$$Av_2 = v_1 + 5v_2,$$

where the second equality reduces to $(A - 5I)v_2 = v_1$. Here, v_2 is called a generalized eigenvector (or a principal vector) for v_1.

Solving $(A - 5I)v_2 = v_1$, we obtain $v_2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$. Then $X = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$.

Note that there is a theorem that states that eigenvectors and principal eigenvectors are linearly independent, so we know that X will be linearly independent.

Definition. We define the right principal eigenvectors of degree k associated with $\lambda \in \Lambda(A)$ is a vector such that

$$(A - \lambda I)^k = 0$$
$$(A - \lambda I)^{k-1} \neq 0.$$

Example. Suppose $A \in \mathbb{C}^{3 \times 3}$, and suppose there is one eigenvalue λ with algebraic multiplicity 3, with only one eigenvector. Then, if we look at

$$A \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} = \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix} \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix},$$

we can get the following equations:

$$Av_1 = \lambda v_1$$
$$Av_2 = v_1 + \lambda v_2$$
$$Av_3 = v_2 + \lambda v_3.$$

This means that v_1 is an eigenvector, $(A - \lambda I)v_2 = v_1$, so v_2 is a generalized eigenvector of degree 2. Lastly, we have $(A - \lambda I)v_3 = v_2$ so v_3 is a generalized eigenvector of degree 3.

Algorithm for computing JCF.

- Find the eigenvalues.

- For each eigenvalues, find the eigenvectors.

- Now:

  - If we have n eigenvectors, we are done.

  - If not: for each linearly independent eigenvector with eigenvalue λ such that $AM(\lambda) \geq GM(\lambda)$, we compute degree 2 generalized eigenvectors.

- If # of eigenvectors + # of generalized eigenvectors of degree $2 = n$, we are done. If the sum is $< n$:

  - Continue: for each generalized eigenvector of degree 2, find (if possible), principle vectors of degree 3.

  - Recurse.

To compute X, just put the generalized eigenvectors in order of increasing degree, and set

$$J = \text{diag}(J_{l_1}(\lambda_1), J_{l_2}(\lambda_2), \ldots, J_{l_k}(\lambda_k)).$$

20 Lecture 26: 8-8-19

Theorem. If we have the system $x' = Ax$, $x(t_0) = x_0$, the solution to the IVP is given by

$$x(t) = e^{A(t-t_0)} \cdot x_0.$$

Lemma. We have that $\frac{d}{dt} e^{At} = Ae^{At} = e^{At} A$.

Proof. (Sketch). Note that

$$\frac{d}{dt}\left[I + At + \frac{A^2t^2}{2!} + \frac{A^3t^3}{3!} + \ldots\right] = A + A^2t + \frac{A^3t^2}{2!} + \frac{A^4t^3}{3!} + \ldots$$

$$= A\left[I + At + \frac{A^2t^2}{2!} + \ldots\right]$$

$$= Ae^{At}.$$

(Note that here, we're assuming that we can differentiate term by term, which is not always true.[2].) □

Now, we will verify the theorem. We know that

$$\frac{dx(t)}{dt} = Ae^{A(t-t_0)} \cdot x_0 = Ax(t)$$

$$X(t_0) = e^{A \cdot 0} \cdot x_0 = x_0.$$

This verifies the conditions of the theorem.

To solve this problem:

- First, compute the JCF of the matrix A.
- Use the JCF to compute e^{At}.
- Use the initial condition to the get the solution of the IVP.

Example. Consider the matrix $A = \begin{pmatrix} -4 & 4 \\ -1 & 0 \end{pmatrix}$, and consider the system

$$x_1' = -4x_1 + 4x_2$$
$$x_2' = -x_1,$$

with initial conditions

$$x_1(2) = 2$$
$$x_2(2) = 4.$$

To solve this, we need to compute the matrix exponential of this. We can compute that

$$A = XJX^{-1}$$

$$X = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}; \qquad J = \begin{pmatrix} -2 & 1 \\ 0 & -2 \end{pmatrix}.$$

And we found that

$$e^{At} = Xe^{Jt}X^{-1} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} e^{-2t} & te^{-2t} \\ 0 & e^{-2t} \end{pmatrix}\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}$$

$$= \begin{pmatrix} e^{-2t} - 2te^{-2t} & 4te^{-2t} \\ -te^{-2t} & e^{-2t} + 2te^{-2t} \end{pmatrix}.$$

And to find the solution, we can compute:

$$x(t) = e^{A(t-t_0)} \cdot x_0,$$

(we will skip this computation for now because it is tedious).

Next week, we will discuss LU decomposition. It's an important idea, but it will not appear on the final.

---

[2]For more on this, see https://www.dpmms.cam.ac.uk/ agk22/uniform.pdf

21   Lecture 27: 8-12-19

The idea of LU decomposition is to solve equations of the form $Ax = b$, where $A \in \mathbb{R}^{m \times n}$, b is a vector, and x is a vector of unknowns.

- **Case 1.** The easiest case is when A is upper triangular.

- **Case 2.** When we have an equation like this:

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 3 & 4 & 5 \\ 1 & 2 & 4 \end{pmatrix},$$

    we can just use elementary operations to reduce this to case 1.

- **Case 3.** Things can a bit more subtle when we have 0s in the coefficients, e.g.

$$A = \begin{pmatrix} 3 & 2 & 7 & 4 \\ -6 & -4 & -14 & -16 \\ 9 & 6 & 23 & 10 \\ 0 & 0 & -2 & -2 \end{pmatrix}.$$

    But we can continue as before, and we end up with

$$A' = \begin{pmatrix} 3 & 2 & 7 & 4 \\ 0 & 0 & 0 & -8 \\ 0 & 0 & 2 & -2 \\ 0 & 0 & -2 & -2 \end{pmatrix}.$$

    The idea is to switch rows and columns to ensure that $A'_{2,2}$ is no longer 0s. In this case, we will interchange R_2 and R_3, and C_2 and C_3. Note that we have to be a bit careful, since this will change the order of the solutions.

Definition. The LU-decomposition of A consists of 2 matrices L and U, where L is an invertible lower triangular matrix, and U is an upper triangular matrix.

Definition. We say that a matrix is k-upper triangular if the first k columns are upper triangular.

Definition. We say that the k-step LU decomposition of A is $A = L_k U_k$, where L_k is an invertible k-lower triangular matrix, and U is a k-upper triangular matrix.

> Add an overview of LU decomposition algorithm

22   Lecture 28: 8-13-19

Recall that given $A \in \mathbb{C}^{m \times n}$, the LU decomposition of A is matrices L and M such that:

- L is a lower triangular matrix with determinant 1 ($L \in \mathbb{C}^{m \times m}$).

- U is an upper triangular matrix with no zero on the diagonal ($U \in \mathbb{C}^{m \times n}$),

- such that $A = LU$.

Definition. We described the r-step LU decomposition were $A = L_r U_r$, where we have .

> fill in block matrices stuffj

Algorithmically, we might have something as follows:

- **(Input.)** $A \in \mathbb{C}^{m \times n}, m, n, r$.

- **(Output.)** L_k, U_k, the k step LU factorization where k is the max number $\leq r$ such that the algorithm can continue.[3]

---

[3]This is necessary since we are doing the algorithm without pivoting.

```
    U = A;
    L = I_m;
    if m = 1 # we are done, nothing to eliminate.
      return L, U
      for k = 1, ..., r do
        (step k of Gaussian elimination)
      for i = k+1, ..., m do
      \dots
```

Source: page 57 of Kazeev's notes.

23   Final review

Definitions.

- Field $(ACID)^2 roughly Vector space (ACID, AID)$

- SVD definition, and existence

- projection (square, ortho)

- pseudoinverse (briefly)

- norms, induced norms

- 7.29 part 1 (defn of p-norms)

- Holder's inequality For complex vectors - we have

$$| < x, y > | \leq ||x||_p ||y||_q,$$

where $\frac{1}{p} + \frac{1}{q} = 1$.

- Unitary invariance of 2-norms. If U is unitary invariant, then $||Ux||_2 = ||x||_2$. Unitary: $UU^H = I$.

- Pythag

- 7.46 - 7.47 (operator matrix norm)

  Operator norm:

$$|| \cdot ||_{1,\infty} = \max_{x \neq 0} \frac{||Ax||_1}{||x||_\infty}$$

$$||A||_{p,q} = \max_{||X||_q \neq 0} \frac{||Ax||_p}{||x||_q}.$$

- First equation, defn. of Frobenius

  It's just L^2 of all entries.

- Least squares problem

- Spectrum (set of all EVs).

- Solution of linear / ODE

  Theorems.

  - 3.19, P1 If A, B in $R^{n \times n}$, then

    – $0 \leq \text{rank}(A + B) \leq \text{rank}(A) + \text{rank}(B)$.

- 4.11 (pseudoinverse under change of basis) $(UAV)^+ = V^T A^+ U^T$. Suppose U, V are orthogonal.

- 5.11 (properties of SVD)

    – Rank A is R, count of SVs.

    – dyadic outer prod.

    – eigenvector relations $Av_i =_i u_i$

    – $R(U_1) = R(A) = N(A^T)^\perp$

  <++>

- 7.5 P is the matrix of an orthogonal projection onto $R(P)$ iff $P^2 = P = P^T$. sym + square.

- 9.12 (spectrum of a Hermitian matrix) Recall hermitian : $A = A^H$. All eigenvalues must be real. Take hermitian TP and conclude $\bar{\lambda} = \lambda$.

- 9.13 (orthogonality of the eigenspaces of a Hermitian) If λ, μ are distinct eigenvalues of A with x, z right EVs, then x, z are orthogonal.

- 9.14 (linear independence of eigenvectors - distinct eigenvalues). If λ_i are distict, then right / left eigenvectors are LI.

brief:

- Row / column ranks are equal. Follows from $N(A)^\perp = R(A^T)$.

- 3.21, P1, 3 (brief) $R(A) = R(AA^T)$. $N(A) = N(A^T A)$.

- 7.12 / 7.18 (brief) $\langle x, y \rangle_Q = x^T Q y$, weighted inner product. Works for complex case.

- 8.1 (linear least squares problem)

  Let $A \in R^{m \times n}, B^{m \times k}$. Then, the gneeral solution to

  $$\min_X ||AX - B||_2$$

  is of the form

  $$X = A^+ B + (I - A^+ A)Y,$$

  where Y is arbitrary. Key idea: if x is full rank, then $x = (A^T A)^{-1} A^T b$.

  It's $(A^T A)^{-1} A^T b$. $(A^T A)^{-1} A^T b$.

- 9.20 Matrix exponential under similarity

  If $X^{-1} A X = \Lambda$, where λ is diag, we have

  $$e^{tA} = X \operatorname{diag}(e^{\lambda_1 t}, \ldots, e^{\lambda_n t}) X^{-1}.$$

- 9.25 determinant of a matrix is product of eigenvalues (counted w multiplicity). [x]

Homework 1.

- 3.1 Note any element A can be writte nas $A = A' + A''$. Not hard. Show that the intersection is triv.

- 3.4. We have $M_n'' = (M_n')^\perp$ and $M_n' = (M_n'')^\perp$. Orthogonality of matrix subspaces. Just show orthogonality.

- 3.6. $R \subseteq S$ iff $S^\perp \subseteq R^\perp$. Makes sense.

- 4.1 did this.

- 4.4 fine.

- 5.3. If A, B have the same right EVs, then $A = CDC^{-1}, B = CEC^{-1}$. Change of basis. Then using the prev expressions and commutativity,

$$AB = (CDC^{-1})(CEC^{-1}) = CDEC^{-1} = CEDC^{-1} = (CEC^{-1})(CDC^{-1}) = BA.$$

  makes sense.

- 5.4. Take conj. transposes - show that $-\lambda = \lambda$.

- 5.5. Diagonalize and go.

- 6.1. Verify axioms - $< A, A > \geq 0$ with 0 iff $A = 0$. Linearity

- 7.1 Use normalize unit vectors / scalar.

- 7.3a. Use conj. transpostition.

- 7.4. Max row sum. Show upper / lower bound. $||A||_2$ is σ_1 is harder. Note: 2-norms are unitarily invariant. $||S||_2 = \sigma_1$. Consider various vectors, then you get something that works.

- 8.1a. just use det = prod of diagonal.

Homework 2.

- 4.1. $AA^T = A^T A = I$. Since $v_i \in \mathbb{R}^n$ are orthonormal, $\langle Av_i, Av_j \rangle = v_i^T A^T Av_j$, then use A is orthogonal and get orthogonality.

  Then ineer product of Av_i, Av_j is $v_i^T A^T Av_j$, 0 if $i \neq j$, 1 otherwise. Since $v_i^T(A^T Av_j) = 1$. The v_i are n orthonormal vectors in $\mathbb{R}^n$, so they form a basis.

- 4.4. Compute lim / PE. Conditions (MP): $ABA = A, BAB = B, (AB)^T = AB, (BA)^T = BA$.

- 5.1. CH to compute A^{-1}. Multiply by A^{-1}

- 5.2. Char poly of A is $\det(A - xI) = x^3 + 5x^2 - 8x + 4$. Consider each eigenspace / basis. Defective - does not have a complete basis of EVs (not diagonalizable).

- 6.2 ∞ and 2 is easy. $\infty \leq 2$, but $v = [1, 0, \dots]$ And $2^2 \leq n\infty^2$. Makes sense.

- 6.3 Need to show $P^2 = P = P^T$ (symmetric and $P^2 = P$, maybe idempotent? not sure).

- 7.5. matrix norms of $2, \infty$ are equiv.

maybe just recall vector case, and try to generalize

- 8.3 See below.

- 8.4 JNF. The possible JNF blocks correspond to the partitions. Consider diag matrices with 1 on superd. And then just partition stuff.

  Don't need to compute eigenvectors to get JNF.