Note: This document is mainly on Chapter 4, 5, 6, and SVD & LU

For Chapter 1 & 2 and most of Chapter 3, you may review them by yourself.

Note: Don't spend too much time on this note, The best way to prepare for the Final is to do all the practice exams.

# Basic Notations

$$\sum_{i=1}^{n} a_i = a_1 + a_2 + \cdots + a_n$$

"#": "the number of"

"$\forall$": "for all ..."

"$\exists$": "there exist ..."

"Def": "definition"

"Prop": "Property / Proposition"

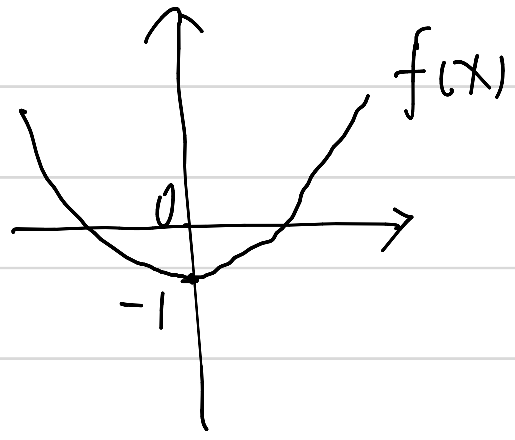
"Thm": "Theorem"

"$\arg\min_{x} f(x)$": minimizer of $f(x)$

# Explanation on $\underset{x}{\operatorname{argmin}} f(x)$ :

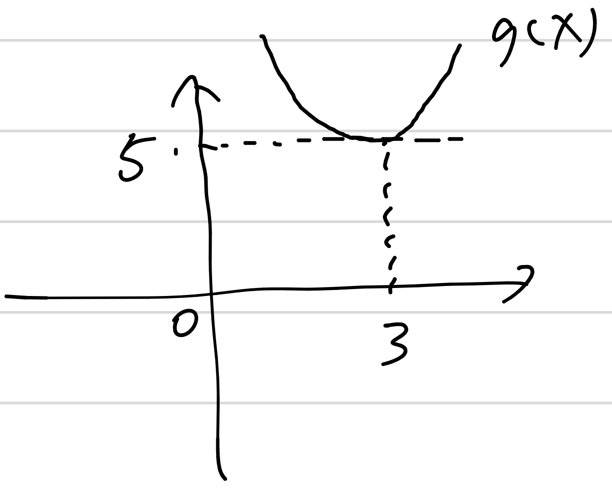
$\underset{x}{\operatorname{argmin}} f(x)$ means " the x that minimizes f(x) "

Example 1: for $f(x) = x^2 - 1$ :




f(x) reaches its minimum value at $x=0$
and that value is $-1$
Then we say $\underset{x}{\min}\ x^2-1 = -1$ , $\underset{x}{\operatorname{argmin}}\ x^2-1 = 0$

Example 2: for $g(x) = (x-3)^2 + 5$ :




g(x) reaches its minimum value at $x=3$
and that value is $5$
Then we say $\underset{x}{\min}\ (x-3)^2+5 = 5$ , $\underset{x}{\operatorname{argmin}}\ (x-3)^2+5 = 3$

Some useful $2\times2$ matrices for finding counterexamples for True / False questions

① $A=\begin{bmatrix}1&0\\0&1\end{bmatrix}$ : $\forall x\in\mathbb{R}^2$, $Ax=x$

② $A=\begin{bmatrix}1&0\\0&0\end{bmatrix}$ : $A^2=A$, not invertible, has two distinct eigenvalues

③ $A=\begin{bmatrix}0&1\\0&0\end{bmatrix}$ : $A^2=\begin{bmatrix}0&0\\0&0\end{bmatrix}$ eigenvalues are both $0$

④ $A=\begin{bmatrix}\pm1&0\\0&\pm1\end{bmatrix}$ : $A^{-1}=A$

Important concept on Chapter 3:

4 ways to Compute Matrix Multiplication:

Let $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$ $\Rightarrow AB \in \mathbb{R}^{m \times p}$

① : Basic formula

Denote $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$, $B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & \cdots & \cdots & b_{2p} \\ \vdots & & & \\ b_{n1} & \cdots & \cdots & b_{np} \end{bmatrix}$

Then let $C = AB = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ c_{m1} & \cdots & & c_{mp} \end{bmatrix}$

then $c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} = a_{i1} b_{1j} + a_{i2} b_{2j} + \cdots + a_{in} b_{nj}$

If we denote $A = \begin{bmatrix} — & \vec{r}_1^{\,T} & — \\ — & r_2^T & — \\ & \vdots & \\ — & r_m^T & — \end{bmatrix}$, $B = \begin{bmatrix} | & & | \\ \vec{b}_1 & \cdots & \vec{b}_p \\ | & & | \end{bmatrix}$

then $c_{ij} = \vec{r}_i^{\,T} \vec{b}_j = \vec{r}_i \cdot \vec{b}_j$

Example:

$$\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \begin{bmatrix} x & y \\ z & w \\ u & v \end{bmatrix} = \begin{bmatrix} ax+bz+cu & ay+bw+cv \\ dx+ez+fu & dy+ew+fv \end{bmatrix}$$

$2 \times 3$ $\quad$ $3 \times 2$ $\quad$ $2 \times 2$

② By Column:

Denote $A = \begin{bmatrix} | & & | \\ \vec{a}_1 & \cdots & \vec{a}_n \\ | & & | \end{bmatrix}$, $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$

then $A\vec{x} = \sum_{i=1}^{n} x_i \vec{a}_i = x_1\vec{a}_1 + x_2\vec{a}_2 + \cdots + x_n\vec{a}_n$

Let $B = \begin{bmatrix} | & & | \\ \vec{b}_1 & \cdots & \vec{b}_p \\ | & & | \end{bmatrix} \in \mathbb{R}^{n\times p}$

Then $AB = \begin{bmatrix} | & | & & | \\ A\vec{b}_1 & A\vec{b}_2 & \cdots & A\vec{b}_p \\ | & | & & | \end{bmatrix}$

Example: $\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \begin{bmatrix} \pi & 10 \\ -7 & e \end{bmatrix} = \begin{bmatrix} | & | \\ \vec{v}_1 & \vec{v}_2 \\ | & | \end{bmatrix}$

where $\vec{v}_1 = \pi \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} + (-7) \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$, $\vec{v}_2 = 10 \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} + e \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$

---

Application:

If you have $T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} ax + by + cz \\ rx + ty + sz \end{bmatrix}$

then you can directly get $A = \begin{bmatrix} a & b & c \\ r & t & s \end{bmatrix}$

(standard matrix)

Note: you must first determine the shape of A by looking at the # components of the input & output in our question they are 3 & 2, so A is 2×3

③ By Row: (Same logic as ② taking transpose)

Let $B = \begin{bmatrix} - \vec{s}_1^T - \\ \vdots \\ - \vec{s}_n^T - \end{bmatrix}$, let $\vec{r} = \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix} \in \mathbb{R}^n$

$$\vec{r}^T B = \sum_{i=1}^{n} r_i \vec{s}_i^T = r_1 \vec{s}_1^T + \cdots + r_n \vec{s}_n^T$$

Denote $A = \begin{bmatrix} - \vec{r}_1^T - \\ \vdots \\ - \vec{r}_m^T - \end{bmatrix}$, $AB = \begin{bmatrix} - \vec{r}_1^T B - \\ - \vec{r}_2^T B - \\ \vdots \\ - \vec{r}_m^T B - \end{bmatrix}$

Example & application: row reduction:

Let $A = \begin{bmatrix} 5 & 6 \\ 0 & 2 \end{bmatrix}$

Doing row reduction:

$$\underbrace{\begin{bmatrix} 5 & 6 \\ 0 & 2 \end{bmatrix}}_{A} \xrightarrow[R_2 = R_2]{R_1 = R_1 - 3R_2} \underbrace{\begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix}}_{\text{let's call it } A'} \xrightarrow[R_2 = \frac{1}{2}R_2]{R_1 = \frac{1}{5}R_1} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{I}$$

we can write $\begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix} A = A'$, $\begin{bmatrix} \frac{1}{5} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} A' = I$

we now have $\left( \begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{5} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \right) \cdot A = I$

and that's how we get inverse by row reduction

④ : column × row

Let $A = \begin{bmatrix} | & & | \\ \vec{a}_1 & \cdots & \vec{a}_n \\ | & & | \end{bmatrix}$ , $B = \begin{bmatrix} — & \vec{s}_1^{\,T} & — \\ & \vdots & \\ — & \vec{s}_n^{\,T} & — \end{bmatrix}$

$AB = \sum_{i=1}^{n} \vec{a}_i \vec{s}_i^{\,T}$ , $\underbrace{\text{where each } \vec{a}_i \vec{s}_i^{\,T} \in \mathbb{R}^{m \times p}}$

For example $\underset{3\times 1}{\begin{bmatrix} a \\ b \\ c \end{bmatrix}} \underset{1\times 2}{\begin{bmatrix} x & y \end{bmatrix}} = \underset{3\times 2}{\begin{bmatrix} ax & ay \\ bx & by \\ cx & cy \end{bmatrix}}$

This is complicated but will help us to understand SVD.

# More on Chapter 3:

For linear transformation $T: \mathbb{R}^n \to \mathbb{R}^m$, the domain is $\mathbb{R}^n$, codomain is $\mathbb{R}^m$

The standard matrix is a $m \times n$ matrix given by $\begin{bmatrix} | & & | \\ T(\vec{e}_1) & \cdots & T(\vec{e}_n) \\ | & & | \end{bmatrix}$

where $\{\vec{e}_1, \ldots, \vec{e}_n\}$ is the standard basis of domain

Please review the following concepts by yourself:

Linear Transformation, One-to-One, Onto, Finding Inverse

# Old Chapter 4 & 5 review

## Chapter 4: Determinant

4.1 [definition] — 2 pages

4.2 [cofactor] — 1 page

4.3 [Volumes] — 1 page

## Chapter 5: eigenvalue/vector

5.1 [definitions] — 1 page

5.2 [ $f(\lambda)$ ] — 2 pages

5.4 [diagonalization] — 4 pages

5.5 [complex ] — 4 pages

(For 5.5, only need to know the content of the first two pages)

Basic knowledge:

① In most cases for matrices $A$ and $B$, $AB \neq BA$

If you have matrix $A$ and identity matrix $I$, then $AI = IA = A$

That is, if you have matrices $A, B$ and $C$ and $A = B$, then you have $\begin{cases} CA = CB \\ AC = BC \end{cases}$ but you don't have $AC = CB$.

② Scalar product: for any $A \in \mathbb{R}^{n\times n}$, $\vec{v} \in \mathbb{R}^n$,

$\alpha A = \alpha I A$, $\alpha \vec{v} = \alpha I \vec{v}$, $I$ is the identity matrix of $\mathbb{R}^{n\times n}$

③ If $A\vec{v} = B\vec{v}$ for a certain $\vec{v}$, we cannot say $A = B$

Example: $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & \pi \\ 0 & e \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, but $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \neq \begin{bmatrix} 1 & \pi \\ 0 & e \end{bmatrix}$

☆ However, if $A\vec{v} = B\vec{v}$ for all $\vec{v} \in \mathbb{R}^n$, then $A = B$

Proof:

Denote $A = [\vec{a}_1 \cdots \vec{a}_n]$, $B = [\vec{b}_1 \cdots \vec{b}_n]$. Let $\vec{e}_1 \cdots \vec{e}_n$ be the standard basis of $\mathbb{R}^n$.

Then for all $k \in \{1, \cdots, n\}$ $A\vec{e}_k = B\vec{e}_k \Rightarrow \vec{a}_k = \vec{b}_k$

going over all $\vec{v} = \vec{e}_k$ we have $A = B$

Quiz 7 review:

Q2: For $A \in \mathbb{R}^{n\times n}$, $p \in \mathbb{R}$, then $pA = pIA$

where $pI = \begin{bmatrix} p & & 0 \\ & \ddots & \\ 0 & & p \end{bmatrix} \Rightarrow \det(pI) = p^n$

thus $\det(pA) = p^n \det(A)$

Q3: If $T(x) = Ax$, $U(x) = Bx$,

then $(T \circ U)^{-1}(x) = (AB)^{-1}x = B^{-1}A^{-1}x = U^{-1} \circ T^{-1}(x)$

Regular method:

① Find the standard matrix $A$ and $B$ for $T$ and $U$

② Compute $AB$ and find $(AB)^{-1}$ by row reduction

Alternative Method

① Find $T^{-1}$ and $U^{-1}$ in a literal way:

$T$: counterclockwise rotation by $45^\circ \Longleftrightarrow T^{-1}$: clockwise rotation by $45^\circ$

$\Rightarrow A^{-1} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$

$U$: stretching by $2 \Longleftrightarrow U^{-1}$: stretching by $\frac{1}{2}$

$\Rightarrow B^{-1} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$

Then get the answer $(AB)^{-1} = B^{-1}A^{-1}$

$\boxed{4.1}$ [basics of determinant] - 2 pages

**Essential Definition.** The **determinant** is a function

$$\det\colon \{\text{square matrices}\} \longrightarrow \mathbf{R}$$

satisfying the following properties: ☆ The first 3 properties also applies to columns

1. Doing a row replacement on $A$ does not change $\det(A)$.
2. Scaling a row of $A$ by a scalar $c$ multiplies the determinant by $c$.
3. Swapping two rows of a matrix multiplies the determinant by $-1$.
4. The determinant of the identity matrix $I_n$ is equal to 1.

Notation: $\det(\vec{a}_1, \cdots, \vec{a}_n) = \det[\vec{a}_1 \ \cdots \ \vec{a}_n]$

Formula for the determinant for 2x2 and 3x3 matrix:

$$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

$$\det\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = a(ei-fh) + b(fg-di) + c(dh-ge)$$

Prop: If $A \in \mathbb{R}^{n\times n}$ has an all-zero row (or column) then $\det(A) = 0$

Def: A square matrix $A \in \mathbb{R}^{n\times n}$, $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & \ddots & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{n1} & \cdots & \cdots & a_{nn} \end{bmatrix}$

is called as $\begin{cases} \text{upper-triangular matrix if } a_{ij} = 0 \text{ for } i > j \\ \text{lower-triangular matrix if } a_{ij} = 0 \text{ for } i < j \end{cases}$

diagonal matrix if $a_{ij} = 0$ for $i \neq j$

(diagonal matrix is both upper-triangular and lower-triangular)

Prop: If $A \in \mathbb{R}^{n\times n}$ is upper- or lower-triangular, then

$\det(A) = a_{11} \cdot a_{22} \cdots a_{n-1\,n-1} \cdot a_{nn}$ (the product of diagonal entries)

**Recipe: Computing determinants by row reducing.** Let $A$ be a square matrix. Suppose that you do some number of row operations on $A$ to obtain a matrix $B$ in row echelon form. Then

$$\det(A) = (-1)^r \cdot \frac{\text{(product of the diagonal entries of } B)}{\text{(product of scaling factors used)}},$$

where $r$ is the number of row swaps performed.

**Def!** the transpose of matrix $A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & & a_{nn} \end{bmatrix}$ is $A^T$

where $(A^T)_{ij} = a_{ji}$

**Prop:** $(\alpha A + \beta B)^T = \alpha A^T + \beta B^T$

**Prop:** ① let $A, B \in \mathbb{R}^{n\times n}$, then $\det(AB) = \det(A)\det(B)$

Note: Although we know in most cases $AB \neq BA$, we always have $\det(AB) = \det(A)\det(B) = \det(B)\det(A) = \det(BA)$

② Assume matrix $A$ invertible, $1 = \det(I) = \det(A)\det(A^{-1})$

$\Rightarrow \det(A) = \frac{1}{\det(A^{-1})}$

③ $\det(A^T) = \det(A)$

**Thm** Denote $A = [\vec{a}_1 \cdots \vec{a}_n] \in \mathbb{R}^{n\times n}$, let $\vec{x} \in \mathbb{R}^n$

Denote $A_k = [\vec{a}_1 \cdots \vec{a}_{k-1}\ \vec{x}\ \vec{a}_{k+1} \cdots \vec{a}_n]$ (the k-th column replaced by $\vec{x}$)

Then the transformation $T(\vec{x}) = \det(A_k)$ is linear

i.e

$\det([\vec{a}_1 \cdots \vec{a}_{k-1}\ \alpha\vec{x}+\beta\vec{y}\ \vec{a}_{k+1} \cdots \vec{a}_n])$

$= \alpha \det([\vec{a}_1 \cdots \vec{a}_{k-1}\ \vec{x}\ \vec{a}_{k+1} \cdots \vec{a}_n]) + \beta \det([\vec{a}_1 \cdots \vec{a}_{k-1}\ \vec{y}\ \vec{a}_{k+1} \cdots \vec{a}_n])$

**Remark** (Alternative defining properties). In more theoretical treatments of the topic, where row reduction plays a secondary role, the defining properties of the determinant are often taken to be:

1. The determinant $\det(A)$ is multilinear in the rows of $A$.
2. If $A$ has two identical rows, then $\det(A) = 0$.
3. The determinant of the identity matrix is equal to one.

$\boxed{4.2}$ — 1 page

[Cofactor]

**Definition.** Let $A$ be an $n \times n$ matrix.

1. The $(i,j)$ **minor,** denoted $A_{ij}$, is the $(n-1)\times(n-1)$ matrix obtained from $A$ by deleting the $i$th row and the $j$th column.

2. The $(i,j)$ **cofactor** $C_{ij}$ is defined in terms of the minor by

$$C_{ij} = (-1)^{i+j}\det(A_{ij}).$$

ex. let $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$, $A_{12} = \begin{bmatrix} d & f \\ g & i \end{bmatrix}$ $A_{32} = \begin{bmatrix} a & c \\ d & f \end{bmatrix}$

$C_{12} = (-1)^{1+2}(di - fg)$ $C_{32} = (-1)^{3+2}(af - cd)$

Thm $\det(A) = \sum_{j=1}^{n} a_{ij} C_{ij} = \sum_{i=1}^{n} a_{ij} C_{ij}$

**Summary: methods for computing determinants.** We have several ways of computing determinants:

1. *Special formulas for* $2 \times 2$ *and* $3 \times 3$ *matrices.*

   This is usually the best way to compute the determinant of a small matrix, except for a $3 \times 3$ matrix with several zero entries.

2. *Cofactor expansion.*

   This is usually most efficient when there is a row or column with several zero entries, or if the matrix has unknown entries.

3. *Row and column operations.*

   This is generally the fastest when presented with a large matrix which does not have a row or column with a lot of zeros in it.

Thm:

Generalized method of computing inverse by using det:

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & \ddots & & \vdots \\ \vdots & & \ddots & \\ C_{1n} & \cdots & \cdots & C_{nn} \end{pmatrix}$$ (pay attention to the order)

Thm:

Cramer's rule: For $A\vec{x} = \vec{b}$, denote $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, $A = [\vec{a}_1 \cdots \vec{a}_n]$

denote $A_k = [\vec{a}_1 \cdots \vec{a}_{k-1}\ \vec{b}\ \vec{a}_{k+1} \cdots \vec{a}_n]$ then $x_k = \dfrac{\det(A_k)}{\det(A)}$

4.3 [ Volume ] — 1 page

typo. it should be parallelepiped

Def

**Definition.** The **paralellepiped** determined by $n$ vectors $v_1, v_2, \ldots, v_n$ in $\mathbf{R}^n$ is the subset

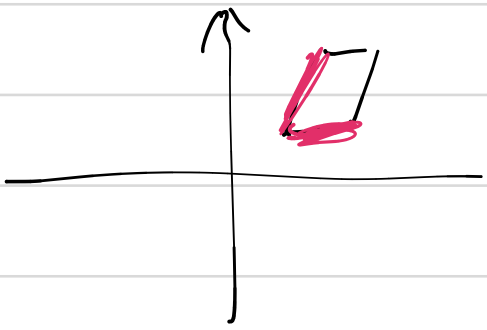
$$P = \{a_1x_1 + a_2x_2 + \cdots + a_nx_n \mid 0 \le a_1, a_2, \ldots, a_n \le 1\}.$$

In other words, a parallelepiped is the set of all linear combinations of $n$ vectors with coefficients in $[0,1]$. We can draw parallelepipeds using the parallelogram law for vector addition.

ex: Consider a parallelogram with vertices $(1,1)$, $(2,3)$, $(3,1)$, $(4,3)$

$(2,3)-(1,1)=(1,2)$

$(3,1)-(1,1)=(2,0)$



We can say it is determined by $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 0 \end{bmatrix}$

☆ The vectors determining the parallelepiped is not unique. In our example, another possibility is $\begin{bmatrix} -1 \\ -2 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 0 \end{bmatrix}$

The number of possibilities is finite.

Thm

**Theorem** (Determinants and volumes). *Let $v_1, v_2, \ldots, v_n$ be vectors in $\mathbf{R}^n$, let $P$ be the parallelepiped determined by these vectors, and let $A$ be the matrix with rows $v_1, v_2, \ldots, v_n$. Then the absolute value of the determinant of $A$ is the volume of $P$:*

$$|\det(A)| = \mathrm{vol}(P).$$

the volume of the parallelogram in our example is $\det\left(\begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix}\right) = 4$

$|\det(A)| \neq \det(|A|)$

Thm

The notation $T(S)$ means the image of the region $S$ under the transformation $T$. In set builder notation, this is the subset

$$T(S) = \{T(x) \mid x \text{ in } S\}.$$

In fact, $T$ scales the volume of *any* region in $\mathbf{R}^n$ by the same factor, even for curvy regions.

This A is the standard matrix of T, not the A above

**Theorem.** *Let $A$ be an $n \times n$ matrix, and let $T : \mathbf{R}^n \to \mathbf{R}^n$ be the associated matrix transformation $T(x) = Ax$. If $S$ is any region in $\mathbf{R}^n$, then*

$$\mathrm{vol}(T(S)) = |\det(A)| \cdot \mathrm{vol}(S).$$

Chapter 5 $\boxed{5.1}$ [Definition] — 1 page

Def:

For square matrix $A \in \mathbb{R}^{n \times n}$, if we find a non-zero $\vec{v} \in \mathbb{R}^n$ such that $A\vec{v} = \lambda \vec{v}$ for a real number $\lambda$, then we say $\vec{v}$ is an eigenvector of $A$, and $\lambda$ as its corresponding eigenvalue.

Example: For $A = \begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ -1 \end{bmatrix}$ are the eigenvectors of $A$ with eigenvalues 2 and 1 respectively.

☆ For each eigenvalue $\lambda$, there are infinitely many vectors to be its eigenvector.

In our example, any vector in the form $\alpha \cdot \begin{bmatrix} 3 \\ -1 \end{bmatrix}$ is an eigenvector of $\begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix}$ with eigenvalue 1.

☆ If $A$ is invertible and $A$'s eigenvalues are $\lambda_1, \cdots, \lambda_n$, then $\frac{1}{\lambda_1}, \cdots, \frac{1}{\lambda_n}$ are the eigenvalue of $A^{-1}$.

# $\boxed{5.2}$ [ Characteristic Polynomial ] – 2 pages

## Formula for finding $\lambda$:

To find the eigenvalues of $A \in \mathbb{R}^{n\times n}$, we observe

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

This means that $\vec{v} \in \text{Nul}(A - \lambda I) \implies A - \lambda I$ not invertible

$\implies$ the eigenvalues of $A$ is all possible $\lambda$ such that $\det(A - \lambda I) = 0$

☆ For each eigenvalue $\lambda$ of $A$, the set of its corresponding eigenvectors is $\text{Nul}(A - \lambda I) \setminus \{\vec{0}\}$

( the null space of $A - \lambda I$ with $0$ excluded )

**def:** The "$\lambda$-eigenspace" is $\text{Nul}(A - \lambda I)$ (which is the set of $\lambda$'s corresponding eigenvectors along with $0$)

☆ $0$ is an eigenvalue of $A$ if and only if $A$ is not invertible.

**Def** $f(\lambda) = \det(A - \lambda I)$ as the characteristic polynomial of $A$

Example $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $f(\lambda) = \det(A - \lambda I) = (a-\lambda)(d-\lambda) - bc$

Def

**Definition.** The **trace** of a square matrix $A$ is the number $\mathrm{Tr}(A)$ obtained by summing the diagonal entries of $A$:

$$\mathrm{Tr}\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2,n-1} & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n-1,1} & a_{n-1,2} & \cdots & a_{n-1,n-1} & a_{n-1,n} \\ a_{n1} & a_{n2} & \cdots & a_{n,n-1} & a_{nn} \end{pmatrix} = a_{11} + a_{22} + \cdots + a_{nn}.$$

**Recipe: The characteristic polynomial of a $2\times 2$ matrix.** When $n=2$, the previous theorem tells us all of the coefficients of the characteristic polynomial:

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

This is generally the fastest way to compute the characteristic polynomial of a $2\times 2$ matrix.

Prop For triangular matrix $A$ with diagonal entries to be $a_{11} \cdots a_{nn}$, $\det(A-\lambda I) = (a_{11}-\lambda)(a_{22}-\lambda)\cdots(a_{nn}-\lambda)$

$\Rightarrow$ its eigenvalues are the diagonal entries.

Prop $A \in \mathbb{R}^{n\times n}$ has n eigenvalues

among them, there are at most n distinct eigenvalues.

This is because we can observe that the characteristic polynomial of a $n\times n$ matrix always has a $(-1)^n\lambda^n$ term

i.e. for $A \in \mathbb{R}^{n\times n}$, $\det(A-\lambda I) = (-1)^n\lambda^n + \cdots\cdots$

Thus $f(\lambda)$ has n roots $\Rightarrow$ A has n eigenvalues.

# 5.4 [Diagonalization] — 4 pages

Def

**Definition.** Let $A$ be an $n \times n$ matrix, and let $\lambda$ be an eigenvalue of $A$.

1. The **algebraic multiplicity** of $\lambda$ is its multiplicity as a root of the characteristic polynomial of $A$.
2. The **geometric multiplicity** of $\lambda$ is the dimension of the $\lambda$-eigenspace.

Example: If you have $\det(A-\lambda I) = (\lambda-17)^3(\lambda+\pi)^2$

then the roots are $\lambda_1=\lambda_2=\lambda_3=17$, $\lambda_4=\lambda_5=-\pi$

We say $\begin{cases} 17 \text{ is an eigenvalue of } A \text{ with algebraic multiplicity } 3 \\ -\pi \text{ is an eigenvalue of } A \text{ with algebraic multiplicity } 2 \end{cases}$

☆ sometimes algebraic multiplicity $\neq$ geometric multiplicity

Example: $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \Rightarrow \lambda^2 = 0 \Rightarrow \lambda_1 = \lambda_2 = 0$

but the 0-eigenspace is $\text{Span}\left\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right\}$ (dimension $= 1$)

☆ The geometric multiplicity of any $\lambda$ is always at least 1.

Def:

**Definition.** An $n\times n$ matrix $A$ is **diagonalizable** if it is similar to a diagonal matrix: that is, if there exists an invertible $n \times n$ matrix $C$ and a diagonal matrix $D$ such that

$$A = CDC^{-1}. \Rightarrow AC = CD$$

Prop If $A = CDC^{-1}$ then $A^k = CD^kC^{-1}$

Assume $A$ is diagonalizable,

★ let $\lambda_1, \dots, \lambda_n$ and $\vec{v}_1, \cdots, \vec{v}_n$ to be $A$'s eigenvalue and eigenvectors.

$$D = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \ \lambda_n \end{bmatrix}, \quad C = [\vec{v}_1 \ \cdots \ \vec{v}_n]$$

(try prove by yourself)

★ here $\vec{v}_1 \cdots \vec{v}_n$ are taken to be the bases of their corresponding eigenspaces

Example: for $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\lambda_1 = \lambda_2 = 1$ (eigenspace $= \mathbb{R}^2$) you can take $\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$; or $\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix}$; but you cannot take $\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \end{bmatrix}$

But for $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $\dim(Nul(A-I)) = 1$ so you can only take $\begin{bmatrix} \alpha \\ 0 \end{bmatrix} \begin{bmatrix} \beta \\ 0 \end{bmatrix}$ for some $\alpha, \beta$

$Nul(A-I) = Span\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}\}$

---

★ We should note that all matrices $A$ can be written as $A \cdot C = C \cdot D$ for $C$ contains its eigenvectors and $D$ being a diagonal matrix.

Example: $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha & \beta \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Therefore, we can say that $A$ is diagonalizable if and only if $C$ can be invertible.

By "can" look at this example:

let $A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \Rightarrow \lambda_1 = \lambda_2 = 2$ $\quad Nul(A-2I) = Span\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}\}$

$V_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $V_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \in Nul(A-2I)$

so you have $A[\vec{v}_1 \ \vec{v}_2] = [\vec{v}_1 \ \vec{v}_2] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$ but $[\vec{v}_1 \ \vec{v}_2]$ is not invertible

However there exists $\vec{v}_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\vec{v}_4 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \in Nul(A-2I)$

such that $A = [\vec{v}_3 \ \vec{v}_4] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} [\vec{v}_3 \ \vec{v}_4]^{-1}$

**Thm** A matrix is diagonalizable if and only if for each eigenvalue $\lambda$, its geometric multiplicity equals to its algebraic multiplicity.

intuition: if not then $\vec{v}_1, \cdots, \vec{v}_n$ not linearly independent

1. *A is diagonalizable.*
2. *The sum of the geometric multiplicities of the eigenvalues of A is equal to n.*
3. *The sum of the algebraic multiplicities of the eigenvalues of A is equal to n, and for each eigenvalue, the geometric multiplicity equals the algebraic multiplicity.*

Prop

This gives us that matrices with distinct eigenvalues are always diagonalizable.

**Recipe: Diagonalization.** Let $A$ be an $n \times n$ matrix. To diagonalize $A$:

1. Find the eigenvalues of $A$ using the characteristic polynomial.
2. For each eigenvalue $\lambda$ of $A$, compute a basis $\mathcal{B}_\lambda$ for the $\lambda$-eigenspace.
3. If there are fewer than $n$ total vectors in all of the eigenspace bases $\mathcal{B}_\lambda$, then the matrix is not diagonalizable.
4. Otherwise, the $n$ vectors $v_1, v_2, \ldots, v_n$ in the eigenspace bases are linearly independent, and $A = CDC^{-1}$ for

$$C = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix} \quad \text{and} \quad D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

where $\lambda_i$ is the eigenvalue for $v_i$.

**Theorem.** *Let A and B be similar $n \times n$ matrices, and let $\lambda$ be an eigenvalue of A and B. Then:*

1. *The algebraic multiplicity of $\lambda$ is the same for A and B.*
2. *The geometric multiplicity of $\lambda$ is the same for A and B.*

exist $C \in \mathbb{R}^{n \times n}$ $\quad A = C \cdot B \cdot C^{-1}$

# 5.5 [Complex eigenvalue] - 4 pages (only need first 2 pages)

**Def:** Imaginary number $i$ : $i^2 = -1$

☆ Any complex number takes the form $z = x + iy$, $x, y \in \mathbb{R}$

☆ For two complex numbers $z_1 = x_1 + iy_1$, $z_2 = x_2 + iy_2$, we have $z_1 = z_2$ if and only if $x_1 = x_2$ and $y_1 = y_2$

**Def:** For $z = x + iy$, we call $Re(z) = x$ (real part of $z$) and $Im(z) = y$ (imaginary part of $z$)

☆ **Basic Properties:** $i^2 = (-i)^2 = -1$, $\frac{1}{i} = -i$

solution to $\lambda^4 = 1$ is $\lambda_1 = 1$, $\lambda_2 = -1$, $\lambda_3 = i$ $\lambda_4 = -i$

Example: For $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, $\det(A - \lambda I) = \lambda^2 + 1 \Rightarrow \begin{cases} \lambda_1 = i \\ \lambda_2 = -i \end{cases}$

$A - \lambda_1 I$:

$$\begin{bmatrix} i & -1 \\ 1 & i \end{bmatrix} \xrightarrow{R_2 = R_2 + iR_1} \begin{bmatrix} i & -1 \\ 0 & 0 \end{bmatrix} \Rightarrow v_2 = v_2,\ iv_1 = v_2 \Rightarrow v_1 = -iv_2$$

$i$-eigenspace $= \text{Span}\left\{ \begin{bmatrix} -i \\ 1 \end{bmatrix} \right\}$

**Def:** The complex conjugate of $z = x + iy$ is $\bar{z} = x - iy$

**Prop:** If $z = x + iy$ is an eigenvalue of $A$, then $\bar{z}$ is, too

Let $\vec{v}$ be $z$'s corresponding eigenvector, then $\overline{\vec{v}}$ is $\bar{z}$'s.

**Def:** Rotation-Scaling matrix: similar to Quiz 7 Q3

define transformations $\begin{cases} T_\theta(x): \text{rotate by } \theta \text{ counterclockwisely} \\ U_t(x): \text{scaling by } t \end{cases}$

Denote $T_\theta(x) = Ax$, $U_t(x) = Bx$, then

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad B = \begin{bmatrix} t & 0 \\ 0 & t \end{bmatrix} = t \cdot I$$

Note: $AB = A \cdot t \cdot I = tIA = BA = \begin{bmatrix} t\cos\theta & -t\sin\theta \\ t\sin\theta & t\cos\theta \end{bmatrix}$

The Rotation-Scaling matrix is $AB$, represented by $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$

where $a = t\cos\theta$, $b = t\sin\theta$

$\Longrightarrow$ $t = \sqrt{a^2+b^2}$, $\theta = \arcsin\left(\frac{b}{t}\right) = \arccos\left(\frac{a}{t}\right) = \arctan\left(\frac{b}{a}\right)$

(Rotating Scaling Theorem next page)

☆ It seems the following will not be tested on the Final, but it is good to know

☆

## Rotating Scaling Theorem!

The illustration on this part from our textbook might be messy. I can explain it in a better way:

Let $A \in \mathbb{R}^{2\times 2}$, with complex eigenvalues $\lambda = x+iy$ $(y \neq 0)$ and $\bar{\lambda}$ let their eigenvector be $u, \bar{u}$

Denote $\vec{u} = \begin{bmatrix} a+ib \\ c+id \end{bmatrix}$, let $\vec{v}_1 = \begin{bmatrix} a \\ c \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} b \\ d \end{bmatrix}$

such that we have $\vec{u} = \vec{v}_1 + i\vec{v}_2$

Then $A(\underbrace{\vec{v}_1 + i\vec{v}_2}_{u}) = \lambda \cdot \vec{u} = (x+iy)(\vec{v}_1 + i\vec{v}_2) = x\vec{v}_1 + iy\vec{v}_1 + xi\vec{v}_2 - y\vec{v}_2$

$= (x\vec{v}_1 - y\vec{v}_2) + i(y\vec{v}_1 + x\vec{v}_2)$

Also $A(\vec{v}_1 + i\vec{v}_2) = (A\vec{v}_1) + i(A\vec{v}_2)$,

Hence $\underbrace{A\vec{v}_1 = x\vec{v}_1 - y\vec{v}_2}$ and $\underbrace{A\vec{v}_2 = y\vec{v}_1 + x\vec{v}_2}$

$A\vec{v}_1 = \begin{bmatrix} | & | \\ \vec{v}_1 & \vec{v}_2 \\ | & | \end{bmatrix} \begin{bmatrix} x \\ -y \end{bmatrix}$ $\qquad$ $A\vec{v}_2 = \begin{bmatrix} | & | \\ \vec{v}_1 & \vec{v}_2 \\ | & | \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix}$

Combining them together,

$$A \begin{bmatrix} | & | \\ \vec{v}_1 & \vec{v}_2 \\ | & | \end{bmatrix} = \begin{bmatrix} | & | \\ \vec{v}_1 & \vec{v}_2 \\ | & | \end{bmatrix} \begin{bmatrix} x & y \\ -y & x \end{bmatrix}$$

Remember how we define them? $\vec{v}_1 = \mathrm{Re}(\vec{u})$, $\vec{v}_2 = \mathrm{Im}(\vec{u})$

$x = \mathrm{Re}(\lambda)$, $y = \mathrm{Im}(\lambda)$

**Block Diagonalization Theorem.** *Let $A$ be a real $n \times n$ matrix. Suppose that for each (real or complex) eigenvalue, the algebraic multiplicity equals the geometric multiplicity. Then $A = CBC^{-1}$, where $B$ and $C$ are as follows:*

- *The matrix $B$ is* **block diagonal**, *where the blocks are $1 \times 1$ blocks containing the real eigenvalues (with their multiplicities), or $2 \times 2$ blocks containing the matrices*
$$\begin{pmatrix} \mathrm{Re}(\lambda) & \mathrm{Im}(\lambda) \\ -\mathrm{Im}(\lambda) & \mathrm{Re}(\Lambda) \end{pmatrix}$$
*for each non-real eigenvalue $\lambda$ (with multiplicity).*

- *The columns of $C$ form bases for the eigenspaces for the real eigenvectors, or come in pairs $\begin{pmatrix} \mathrm{Re}(v) & \mathrm{Im}(v) \end{pmatrix}$ for the non-real eigenvectors.*

Chapter 6: $\boxed{6.1}$ [Definitions] – 2 pages

**Definition.** The **dot product** of two vectors $x, y$ in $\mathbf{R}^n$ is

$$x \cdot y = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1y_1 + x_2y_2 + \cdots + x_ny_n.$$

Thinking of $x, y$ as column vectors, this is the same as $x^T y$.

$\vec{x} \cdot \vec{y} = \vec{x}^T \vec{y} = \vec{y}^T \vec{x}$

☆ We don't always have $A^T B = B^T A$,

$\vec{x}^T \vec{y} = \vec{y}^T \vec{x}$ because it's $1 \times 1$ matrix

☆ $x$ and $y$ must have the same # components, otherwise there is no dot product between them.

**Properties of the Dot Product.** Let $x, y, z$ be vectors in $\mathbf{R}^n$ and let $c$ be a scalar.

1. *Commutativity:* $x \cdot y = y \cdot x$. $x^T y = y^T x$
2. *Distributivity with addition:* $(x + y) \cdot z = x \cdot z + y \cdot z$.
3. *Distributivity with scalar multiplication:* $(cx) \cdot y = c(x \cdot y)$.

The dot product of a vector with itself is an important special case:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1^2 + x_2^2 + \cdots + x_n^2.$$

Therefore, for any vector $x$, we have:

- $\boxed{x \cdot x \geq 0}$
- $\boxed{x \cdot x = 0 \iff x = 0.}$

This leads to a good definition of *length*.

Def: **Fact.** *The* **length** *of a vector $x$ in $\mathbf{R}^n$ is the number*

$$\|x\| = \sqrt{x \cdot x} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}. \geq 0$$

Important Sample Question

**Example.** Suppose that $\|x\| = 2$, $\|y\| = 3$, and $x \cdot y = -4$. What is $\|2x + 3y\|$?

**Solution.** We compute

$$\begin{aligned}\|2x+3y\|^2 &= (2x+3y)\cdot(2x+3y)\\ &= 4x\cdot x + 6x\cdot y + 6x\cdot y + 9y\cdot y\\ &= 4\|x\|^2 + 9\|y\|^2 + 12x\cdot y\\ &= 4\cdot 4 + 9\cdot 9 - 12\cdot 4 = 49.\end{aligned}$$

Hence $\|2x+3y\| = \sqrt{49} = 7$.

☆ Also please review Quiz 10 Question 1

Def x is perpendicular (orthogonal) to y if and only if $\vec{x}\cdot\vec{y} = 0$, written as $\vec{x} \perp \vec{y}$

Thm (Pythagorean) if $\vec{x} \perp \vec{y}$, $\|\vec{x}\|^2 + \|\vec{y}\|^2 = \|\vec{x}-\vec{y}\|^2 = \|\vec{x}+\vec{y}\|^2$

# 6.2 [Orthogonal Complements] — 3 pages

**Def:**

**Definition.** Let $W$ be a subspace of $\mathbf{R}^n$. Its **orthogonal complement** is the subspace

$$W^\perp = \{v \text{ in } \mathbf{R}^n \mid v \cdot w = 0 \text{ for all } w \text{ in } W\}.$$

The symbol $W^\perp$ is sometimes read "$W$ perp."

$n = \dim(W) + \dim(W^\perp)$

**Prop** $(W^\perp)^\perp = W$

**Remark** $W^\perp$ is also a subspace

**Prop:** $(\mathbb{R}^d)^\perp$ in $\mathbb{R}^d$ is $\{\vec{0}\}$

Example: If $W = \text{Span}\left\{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}2\\3\end{pmatrix}\right\} \Rightarrow W = \mathbb{R}^2$

$\dim(W^\perp) = 2 - \underbrace{\dim(W)}_{=2} = 0 \Rightarrow W^\perp = \{\vec{0}\}$

**Thm** Let $\{\vec{w}_1, \cdots, \vec{w}_t\}$ be the basis of $W \subseteq \mathbb{R}^n$

if for vector $\vec{v}$, $\vec{v} \perp w_j$ for all $j = 1, 2, \cdots, t$,

then $v \in W^\perp$

Proof: easy

★ **Thm** $Col(A)^{\perp} = Nul(A^T)$

Proof: let $A = \begin{bmatrix} | & & | \\ \vec{a}_1 & \cdots & \vec{a}_n \\ | & & | \end{bmatrix} \in \mathbb{R}^{m \times n}$

If $W = Col(A)$, then $W^{\perp} = Nul(A^T)$

Intuition: if $\boxed{\vec{v} \in W^{\perp}}$, then $\boxed{\vec{v} \perp Col(A)}$ ( $\vec{v} \perp \vec{a}_k$ for all $k$ )

and $A^T\vec{v} = \begin{bmatrix} - & \vec{a}_1^T & - \\ & \vdots & \\ - & \vec{a}_n^T & - \end{bmatrix} \vec{v} = \begin{bmatrix} \boxed{\vec{a}_1^T \vec{v}} \\ \vdots \\ \vec{a}_n^T \vec{v} \end{bmatrix} = \vec{0}$

Which means that for all $\vec{v} \in W^{\perp}$, $\vec{v} \in Nul(A^T) \implies W^{\perp} \subseteq Nul(A^T)$

It is also easy to show $W^{\perp} \supseteq Nul(A^T)$

Thus $W^{\perp} = Nul(A^T)$ $\quad$ $Col(A) \perp Nul(A^T)$

$\Updownarrow$

$Col(A^T) \perp Nul(A)$

Important Sample Question

**Example.** Compute $W^\perp$, where

Step 1

$$W = \text{Span}\left\{\begin{pmatrix}1\\7\\2\end{pmatrix}, \begin{pmatrix}-2\\3\\1\end{pmatrix}\right\}.$$

Let $A = \begin{pmatrix}1 & -2\\7 & 3\\2 & 1\end{pmatrix}$

$W = Col(A)$

**Solution.** According to the proposition, we need to compute the null space of the matrix

Step 2: $$A^T = \begin{pmatrix}1 & 7 & 2\\-2 & 3 & 1\end{pmatrix} \xrightarrow{\text{RREF}} \begin{pmatrix}1 & 0 & -1/17\\0 & 1 & 5/17\end{pmatrix}.$$

The free variable is $x_3$, so the parametric form of the solution set is $x_1 = x_3/17$, $x_2 = -5x_3/17$, and the parametric vector form is

Step 3: $$\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = x_3\begin{pmatrix}1/17\\-5/17\\1\end{pmatrix}.$$

$$\underbrace{(Col(A))^\perp}_{W^\perp} = Nul(A^T)$$

Scaling by a factor of 17, we see that

$$\boxed{W^\perp} = \text{Span}\left\{\begin{pmatrix}1\\-5\\17\end{pmatrix}\right\}.$$

We can check our work:

Recipe:

$$\text{Span}\{\vec{v}_1, \cdots, \vec{v}_n\}^\perp = Nul\left(\begin{bmatrix} - \vec{v}_1^T - \\ \vdots \\ - \vec{v}_n^T - \end{bmatrix}\right)$$

$$Col(A)^\perp = Nul(A^T) \qquad (Nul(A^T)^\perp = Col(A))$$

$$Col(A^T)^\perp = Nul(A) \qquad (Nul(A)^\perp = Col(A^T)^\perp)$$

Remember $dim(Col(A)) = dim(Col(A^\perp))$

# Section 6.3 [Orthogonal projection] — 2 pages

## Basic Definition:

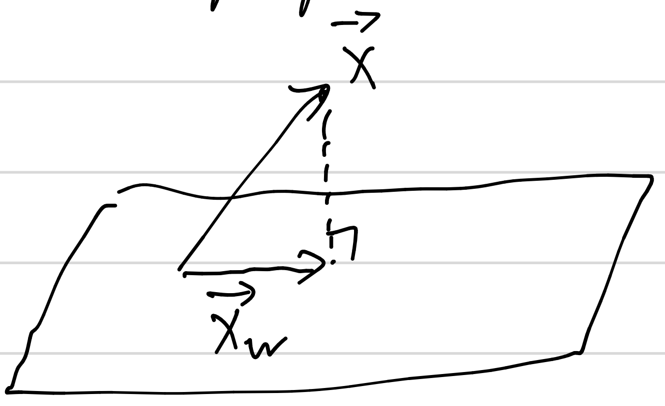
We define $x_W$ to be the orthogonal projection of $\vec{x}$ onto the subspace $W$.

$$\boxed{\text{Formal definition: } \vec{x}_W = \underset{\vec{v} \in W}{\operatorname{argmin}} \| \vec{x} - \vec{v} \|}$$

If you don't know what is "argmin", go to the third page of this document.

☆ **Prop** $(\vec{x} - \vec{x}_W) \perp W$

Intuition: the shortest distance from a point to a plane is the perpendicular distance




and we define $\vec{x}_{W^\perp} = \vec{x} - \vec{x}_W$

☆ : If $\vec{x} \in W^\perp$, then $\vec{x}_W = 0$, $\vec{x} \cdot \vec{x}_W = 0$

☆ : If $\vec{x} \notin W^\perp$, then $\vec{x}_W \neq 0$ $\vec{x} \cdot \vec{x}_W \neq 0$

Finding Orthogonal projection:

Let $W = \text{span}\{\vec{a}_1, \dots, \vec{a}_n\}$, want to find $\boxed{\vec{x}_W}$

Step 1: Let $A = \begin{bmatrix} | & & | \\ \vec{a}_1 & \cdots & \vec{a}_n \\ | & & | \end{bmatrix}$, $A^T = \begin{bmatrix} - & \vec{a}_1^T & - \\ & \vdots & \\ - & \vec{a}_n^T & - \end{bmatrix}$

Step 2: calculate $\underbrace{A^T A}_{=B}$ and $\underbrace{A^T \vec{x}}_{=\vec{v}}$

Step 3: solve $\boxed{A^T A \vec{c} = A^T \vec{x}}$, $\vec{c}$

$B\vec{c} = \vec{v}$

Step 4: (Don't forget !!!) Calculate $\vec{x}_W = A\vec{c}$

Alternative Method:

$A^T A \vec{c} = A^T \vec{x} \Rightarrow \vec{c} = (A^T A)^{-1} A^T \vec{x} \Rightarrow \vec{x}_W = A\vec{c} = A(A^T A)^{-1} A^T x$

we call $A(A^T A)^{-1} A^T$ to be the projection matrix for $\text{Col}(A)$

* Calculating $(A^T A)^{-1}$ is expensive!

Def !

$A(A^T A)^{-1} A^T$ is the projection matrix for $W = \text{Col}(A)$

**Properties of Orthogonal Projections.** *Let $W$ be a subspace of $\mathbf{R}^n$, and define $T: \mathbf{R}^n \to \mathbf{R}^n$ by $T(x) = x_W$. Then:*

1. *$T$ is a linear transformation.*
2. *$T(x) = x$ if and only if $x$ is in $W$.*
3. *$T(x) = 0$ if and only if $x$ is in $W^\perp$.*
4. *$T \circ T = T$.*
5. *The range of $T$ is $W$.*

$\boxed{6.5}$ [ Least Square ] — 1.5 pages

If you have dataset $(x_1, y_1), (x_2, y_2), \cdots, (x_n\ y_n)$ and you want to fit in a certain function:

The least square formula depends on your function:

(1) if you want $y = a + bx$ :

Let $A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$, $\vec{t} = \begin{bmatrix} a \\ b \end{bmatrix}$, $\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$

and solve $A^TA\vec{t} = A^T\vec{y}$ for $a, b$

(2) if you want $y = a + bx + cx^2$ :

Let $A = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}$, $\vec{t} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$, $\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$

and solve $A^TA\vec{t} = A^T\vec{y}$ for $a, b, c$.

and so on for polynomials

(3): if you want $y = a \cdot e^{bx}$, convert it to $\ln y = \ln a + bx$ [ Let's call $\ln a = a'$

Let $A = \begin{bmatrix} 1 & x \\ 1 & x \\ \vdots & \vdots \\ 1 & x \end{bmatrix}$, $\vec{t} = \begin{bmatrix} a' \\ b \end{bmatrix}$, $\vec{y} = \begin{bmatrix} \ln y_1 \\ \vdots \\ \ln y_n \end{bmatrix}$

Then solve $A^TA\vec{t} = A^T\vec{y}$ for $a'$ and $b$, then get $a = e^{a'}$

After you solved $\vec{t}$, you have:

$A\vec{t}$ is the vector that containes the y predicted from our fitted function.

# $\boxed{\text{SVD}}$ — 7 pages

[This is a thorough overview of SVD. If you just want to know the computation, jump to page 42]

## Basic Knowledge:

**Def** We say a square matrix $M$ is symmetric if $M^T = M$

**Thm** If $M \in \mathbb{R}^{n\times n}$ is symmetric, you can find a set of eigenvectors $\{\vec{v}_1, \cdots, \vec{v}_n\}$ of $M$ such that $\vec{v}_i \perp \vec{v}_j$ for any $i \neq j$.
(pairwisely orthogonal)

That is, $$\vec{v}_i^T \vec{v}_j = \vec{v}_i \cdot \vec{v}_j = \begin{cases} \|\vec{v}_i\|^2 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

☆ Since eigenvectors are subject to scaling, we can scale these vectors such that their lengths are all $=1$.
Then, $$\vec{v}_i^T \vec{v}_j = \vec{v}_i \cdot \vec{v}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$
and we call $\{\vec{v}_1, \cdots, \vec{v}_n\}$ an orthonormal basis

**Conclusion:** If matrix $M \in \mathbb{R}^{n\times n}$ is symmetric, then we can find some of its eigen vectors $\vec{v}_1, \cdots, \vec{v}_n$, such that $\{\vec{v}_1, \cdots, \vec{v}_n\}$ forms an orthonormal basis of $\mathbb{R}^n$

For all matrix $A \in \mathbb{R}^{m\times n}$:

① $A^TA \in \mathbb{R}^{n\times n}$ is a square matrix. (obviously)

② Furthermore, $A^TA$ is a symmetric matrix $\left((A^TA)^T = A^TA\right)$

③ From ② we know that we can find an orthonormal basis of $\mathbb{R}^n$ from $A^TA$'s eigenvectors

④ For all $\vec{x} \in \mathbb{R}^n$, $\vec{x}^T A^T A \vec{x} \geq 0$ $\left(\vec{x}^T A^T A \vec{x} = (A\vec{x}) \cdot (A\vec{x}) = \|A\vec{x}\|^2 \geq 0\right)$

(We say a symmetric matrix $M \in \mathbb{R}^{n\times n}$ is "positive semi-definite" if $\forall \vec{x} \in \mathbb{R}^n$, $\vec{x}^T M \vec{x} \geq 0$)

Thus we say $A^TA$ is positive semi-definite

⑤ From ④ we know all eigenvalues of $A^TA$ are non-negative

Proof: Assume $\lambda, \vec{v}$ are a pair of eigenvalue/vector of $A^TA$ which means $A^TA\vec{v} = \lambda \vec{v}$

By ④, $0 \leq \vec{v}^T A^T A \vec{v} = \vec{v}^T \lambda \vec{v} = \lambda \|\vec{v}\|^2$

since $\|\vec{v}\|^2 \geq 0$, we must have $\lambda \geq 0$

⑥: $AA^T \in \mathbb{R}^{m\times m}$ has the same properties above, just changed "n" to "m"

Main part of SVD:

**Def:** For all matrix $A \in \mathbb{R}^{m \times n}$, it has a singular-value decomposition (SVD):

$$A = U \Sigma V^T, \quad \text{where} \quad \begin{matrix} U^{-1} = U^T \\ V^{-1} = V^T \end{matrix}$$

( full SVD )

where:

① $$\Sigma = \begin{cases} \begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ 0 & & \ddots & \sigma_n \\ 0 & \cdots & & 0 \\ \vdots & & & \\ 0 & \cdots & & 0 \end{bmatrix} \in \mathbb{R}^{m \times n} & \text{if } m > n \\ \begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ 0 & & \ddots & \\ & & & \sigma_m \end{bmatrix} & \text{if } m = n \\ \begin{bmatrix} \sigma_1 & & 0 & & 0 & \cdots & 0 \\ & \sigma_2 & & & \vdots & \ddots & \vdots \\ 0 & & \ddots & & \vdots & & \vdots \\ & & & \sigma_m & 0 & \cdots & 0 \end{bmatrix} & \text{if } m < n \end{cases}$$

where $(\Sigma)_{ij} = 0$ if $i \neq j$ and $\sigma_1, \sigma_2, \sigma_3, \cdots$ are $\geq 0$

☆ The diagonal entries can be denoted as $\sigma_1, \sigma_2, \cdots, \sigma_{\min(m,n)}$

For convenience and generality, let's consider the case when $\boxed{m > n}$

For convenience, let's order $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \cdots \geq 0$

(2) $U = \begin{bmatrix} | & & | \\ \vec{u}_1 & \cdots & \vec{u}_m \\ | & & | \end{bmatrix} \in \mathbb{R}^{m\times m}$

where $\{\vec{u}_1, \cdots, \vec{u}_m\}$ are the eigenvectors of $AA^T$ that form an orthonormal basis of $\mathbb{R}^m$.

☆ Their corresponding eigenvalues are $\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2, 0, \cdots, 0$

(if $m \le n$ then they are $\sigma_1^2, \cdots, \sigma_m^2$)

Proof: $AA^T = U\Sigma V^T V \Sigma^T U^T = U\Sigma\Sigma^T U^T$

$\Longrightarrow (AA^T)U = U(\Sigma\Sigma^T)$

Prop: $U^{-1} = U^T$

(3) Let $V = \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{bmatrix} \in \mathbb{R}^{n\times n}$

where $\{\vec{v}_1, \cdots, \vec{v}_n\}$ are the eigenvectors of $A^TA$ that form an orthonormal basis of $\mathbb{R}^n$.

☆ Their corresponding eigenvalues are $\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2$

(if $m \le n$ then they are $\sigma_1^2, \cdots, \sigma_m^2, 0, \cdots, 0$)

Proof: $A^TA = V\Sigma^T U^T U \Sigma V^T = V\Sigma^T\Sigma V^T$

$\Longrightarrow (A^TA)V = V(\Sigma^T\Sigma)$

Prop: $V^{-1} = V^T$

---

☆ Thus we have $AV = U\Sigma$

which means $A\vec{v}_i = \sigma_i \vec{u}_i$ for all $i \in \{1, \cdots, \min(m,n)\}$

Important Observation:

Still consider $\boxed{m > n}$ (which means $\min(m,n) = n$ )

$$U \Sigma V^T = \underbrace{\begin{bmatrix} | & & | \\ \vec{u}_1 & \cdots & \vec{u}_m \\ | & & | \end{bmatrix}}_{m \times m} \underbrace{\begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ 0 & & \ddots & \\ & & & \sigma_n \\ 0 & \cdots & & 0 \\ \vdots & & & \\ 0 & \cdots & & 0 \end{bmatrix}}_{m \times n} \underbrace{\begin{bmatrix} - & \vec{v}_1^T & - \\ & \vdots & \\ - & \vec{v}_n^T & - \end{bmatrix}}_{n \times n}$$

By formula of matrix multiplication,

$$U \Sigma = \underbrace{\begin{bmatrix} | & & | \\ \sigma_1 \vec{u}_1 & \cdots & \sigma_n \vec{u}_n \\ | & & | \end{bmatrix}}_{m \times n}$$

Thus $$U \Sigma V^T = \begin{bmatrix} | & & | \\ \sigma_1 \vec{u}_1 & \cdots & \sigma_n \vec{u}_n \\ | & & | \end{bmatrix} \begin{bmatrix} - & \vec{v}_1^T & - \\ & \vdots & \\ - & \vec{v}_n^T & - \end{bmatrix}$$

By the 4th way of matrix multiplication, on page 8 of this document, we have

$$U \Sigma V^T = \sum_{k=1}^{n} \sigma_k \vec{u}_k \vec{v}_k^T \quad \left( \text{where each } \underbrace{\vec{u}_k}_{m \times 1} \underbrace{\vec{v}_k^T}_{1 \times n} \in \mathbb{R}^{m \times n} \right)$$

☆ Let's denote $r$ to be the number of non-zero $\sigma$

(i.e. denote $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = 0 = \sigma_{r+2} = \cdots = \sigma_{\min(m,n)}$)

$r \leq \min(m,n)$

We are considering the case $m > n$, so $r \leq n$

Then,

$$U \Sigma V^T = \sum_{k=1}^{n} \sigma_k \vec{u}_k \vec{v}_k^T = \sum_{k=1}^{r} \sigma_k \vec{u}_k \vec{v}_k^T + \sum_{k=r+1}^{n} 0 \cdot \vec{u}_k \vec{v}_k^T$$

$$= \sum_{k=1}^{r} \sigma_k \vec{u}_k \vec{v}_k^T$$

(turn it back to matrix):

$$= \underbrace{\begin{bmatrix} | & & | \\ \vec{u}_1 & \cdots & \vec{u}_r \\ | & & | \end{bmatrix}}_{\tilde{U} \in \mathbb{R}^{m \times r}} \underbrace{\begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_r \end{bmatrix}}_{\tilde{\Sigma} \in \mathbb{R}^{r \times r}} \underbrace{\begin{bmatrix} - & \vec{v}_1^T & - \\ & \vdots & \\ - & \vec{v}_r^T & - \end{bmatrix}}_{\tilde{V} \in \mathbb{R}^{r \times n}}$$

(Reduced SVD)

If $m < n$, by similar logic we can still get this.

If some question want you to compute the SVD of a matrix

① If it is not a square matrix, then compute reduced SVD

② If it is a square matrix, then if you want, you can compute the full SVD.

Computing SVD of $A \in \mathbb{R}^{m \times n}$

Step 1: Compute $S = A^T A$

Step 2: Compute the eigenvalues $\lambda_1, \cdots, \lambda_n$ of $A^T A$ by using characteristic polynomial. [They should all $\geq 0$]

Step 2.5 take the square roots of the non-zero eigenvalues as $\sigma_1 = \sqrt{\lambda_1}, \cdots, \sigma_r = \sqrt{\lambda_r}$

Step 3: Find the corresponding eigenvectors of $\lambda_1, \cdots, \lambda_r$, (☆ not $\sigma_1, \cdots, \sigma_r$ !)

which are $\vec{v}_1, \cdots, \vec{v}_r$

and scale $\vec{v}_1, \cdots, \vec{v}_r$ to make their lengths $= 1$

Step 4: Compute $\vec{u}_1 = \frac{1}{\sigma_1} A \vec{v}_1, \cdots, \vec{u}_r = \frac{1}{\sigma_r} A \vec{v}_r$

make sure $\vec{u}_1, \cdots, \vec{u}_r$ all have length $= 1$

Step 5:

write $A = \begin{bmatrix} | & & | \\ \vec{u}_1 & \cdots & \vec{u}_r \\ | & & | \end{bmatrix} \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \end{bmatrix} \begin{bmatrix} - & \vec{v}_1^T & - \\ & \vdots & \\ - & v_r^T & - \end{bmatrix}$

[If $A$ is a square matrix you can also keep those zero-$\sigma$]

# $\boxed{LU}$ — 2 pages

**Basic Knowledge:** Row Reduction can be represented as matrix multiplication.

(If you don't know this, go to page 7)

## Computation of $LU$ of $A \in \mathbb{R}^{n \times n}$ (square matrix)

It is just doing row reduction that:

① cannot swap the position of rows

② cannot scale any row

(① & ② means that we can only do row replacement)

③ when you're going to do row replacement on a certain row, you cannot use the row below it

In this way row reduction can be represented as

$$A \to E_1 A \to E_2 E_1 A \to \cdots \to \underbrace{E_n E_{n-1} \cdots E_2 E_1 A}_{A_{REF}}$$

Note! Because of ②, we just get REF, not RREF

where all $E_k$ are lower-triangular and invertible

$A_{RREF}$ is upper-triangular

Then let $L = E_1^{-1} E_2^{-1} \cdots E_{n-1}^{-1} E_n^{-1}$, $U = A_{REF}$

we have $A = LU$

★ Properties of each $E_i$:

They are all in this form $\begin{bmatrix} 1 & & & \\ & 1 & & \\ a & & \ddots & \\ & & & 1 \end{bmatrix}$

where $\begin{cases} \text{diagonal entries are all } 1 \\ \text{one of the lower-left entries is } a \neq 0 \\ \text{other entries are zero} \end{cases}$

and the inverse is $\begin{bmatrix} 1 & & & \\ & 1 & & \\ -a & & \ddots & \\ & & & 1 \end{bmatrix}$

( this "$-a$" is at the same position as $a$ )

★ Property of $L$:

Its diagonal entries are 1

Application:

When solving $A\vec{x}=\vec{b}$, we have $L(U\vec{x})=\vec{b}$

you can first solve $\vec{y}$ for $L\vec{y}=\vec{b}$, then $\vec{x}$ for $U\vec{x}=\vec{y}$

Thm:

A has LU-decomposition if:

① A is invertible

② all of its "leading matrices" are invertible

k-th-Leading matrices: the upper-left $k\times k$ block of A

example: $A=\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$

1st-leading matrix of A: $[a]$

2nd-leading matrix of A: $\begin{bmatrix} a & b \\ d & e \end{bmatrix}$

proof: very hard