Introduction to Linear Algebra
- Gilbert Strang

10

Singular Value Decomposition

sooraj351729@gmail.com +91-9400635788

# INDEX

(10)

NAME: SOORAJ. S. STD.: SEC.: ROLL NO.: SUB.:

| S. No. | Date | Title | Page No. | Teacher's Sign / Remarks |
|---|---|---|---|---|
| | | INTRODUCTION TO | | |
| | | LINEAR ALGEBRA | | |
| | | - Gilbert Strang, MIT | | |
| | | (5th Edition) | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

$\boxed{7\cdot 2(B)}$ Show that $\sigma_1 \geq |\lambda|_{max}$. The largest singular value dominates all eigenvalues.

Ans: $A = U\Sigma V^T$

$\|Qx\| = \|x\|$ $\Rightarrow$ This applies to $U$ & $V^T$.

since $Q$ is orthogonal matrix

$$\|Ax\| = \|U\Sigma V^T x\| = \|\Sigma V^T x\| \leq \sigma_1 \|V^T x\| = \sigma_1 \|x\|$$

An eigenvector has $\|Ax\| = |\lambda| \|x\|$

$$\therefore \quad |\lambda| \|x\| \leq \sigma_1 \|x\| \Longrightarrow |\lambda| \leq \sigma_1$$

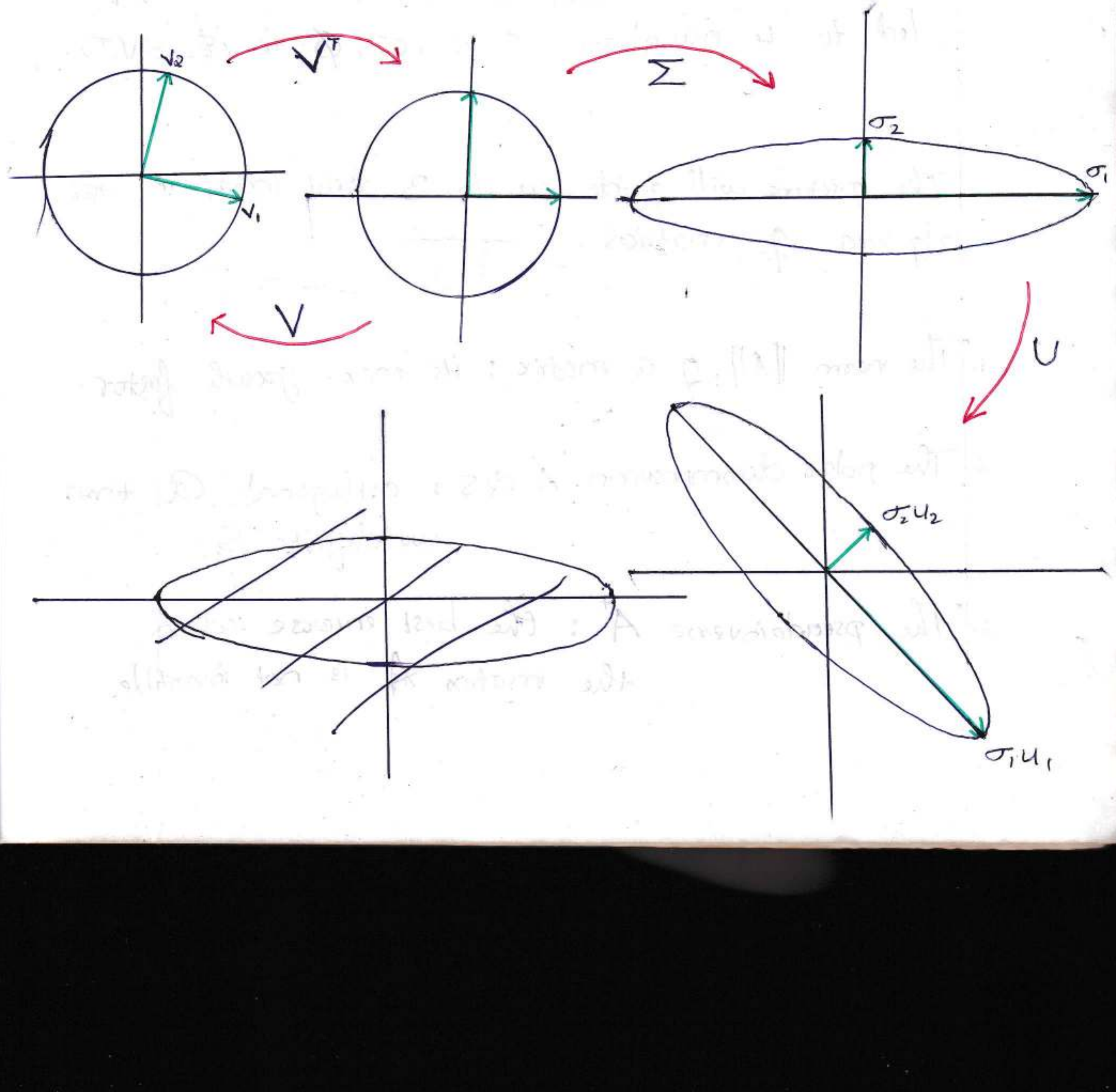
Max. of $\frac{\|Ax\|}{\|x\|}$ is $\sigma_1$.

# The Geometry of the SVD

SVD: $A = U\Sigma V^T = (\text{orthogonal}) \times (\text{diagonal}) \times (\text{orthogonal})$

$= (\text{rotation}) \times (\text{stretching}) \times (\text{rotation})$

$U\Sigma V^T x$ starts with the rotation to $V^T x$. Then $\Sigma$ stretches that vector to $\Sigma V^T x$, and $U$ rotates to $Ax = U\Sigma V^T x$.



This picture applies to a $2 \times 2$ matrix. And not every ~~rotation~~ $2 \times 2$ matrix, because $U$ and $V$ didn't allow for a reflection — all 3 matrices have determinant $> 0$.

This $A$ would have to be invertible because the 3 steps are shown as invertible.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \sigma_1 & \\ & \sigma_2 \end{bmatrix} \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} = U \Sigma V^T$$

$\Rightarrow$ The 4 numbers $a, b, c, d$ in the matrix $A$ led to 4 numbers $\theta, \sigma_1, \sigma_2, \phi$ in its SVD.

This picture will guide us to 3 neat ideas in the algebra of matrices:

1. The norm $\|A\|$ of a matrix: its max. growth factor
2. The polar decomposition, $A = QS$: orthogonal $Q$ times +ve definite $S$.
3. The pseudoinverse $A^+$: the best inverse when the matrix $A$ is not invertible.

## □ The Norm of a Matrix

- $\sigma_1$ is the largest growth factor of any vector $x$.

This largest singular value $\sigma_1$ is the norm of the matrix $A$.

$$\boxed{\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \sigma_1}$$

$$\|x\| = |x_1|^2 + |x_2|^2 + \cdots + |x_n|^2 .$$

- The matrix norm comes from this vector norm when $x = v_1$ and $Ax = \sigma_1 u_1$ and

$$\frac{\|Ax\|}{\|x\|} = \sigma_1 = \text{largest ratio} = \|A\|.$$

$$\|A+B\| \le \|A\| + \|B\| \leftarrow \text{Triangle inequality}$$

$$\|AB\| \le \|A\| \cdot \|B\| \leftarrow \text{Product inequality}$$

Proof $\|A\| = \max_{x \neq 0} \dfrac{\|Ax\|}{\|x\|} = \sigma_1$

$$\Rightarrow \|A\| \ge \frac{\|Ax\|}{\|x\|}$$

$\boxed{\|Ax\| \le \|A\|\|x\|}$ for every vector $x$

Triangle inequality for vectors,

$$\|(A+B)x\| = \|Ax+Bx\| \leq \|Ax\| + \|Bx\| \leq \|A\|\|x\| + \|B\|\|x\|$$

$\div$ing by $|x|$, Take the max. over all $x$.

$$\max_{x\neq 0} \frac{\|(A+B)x\|}{\|x\|} \leq \max_{x\neq 0}\left(\|A\| + \|B\|\right)$$

$$\underline{\underline{\|A+B\| \leq \|A\|+\|B\|}}$$

$$\|ABx\| \leq \|A\|\|Bx\| \leq \|A\|\|B\|\|x\|$$

$\div$ing by $|x|$ & take the maximum over all $x$.

$$\max_{x\neq 0} \frac{\|ABx\|}{\|x\|} \leq \max_{x\neq 0} \|A\|\|B\|$$

$$\underline{\underline{\|AB\| \leq \|A\|\|B\|}}$$

## Norm - idea

Given a vector space $V$,

then a norm, denoted by $\|x\|$ for $x \in V$, is a real # such that

$$\|x\| > 0 \quad , \forall x \neq 0$$

$$\|\alpha x\| = |\alpha| \, \|x\| \quad , \alpha \in \mathbb{R}$$

$$\|x+y\| \leq \|x\| + \|y\|$$

The norm is a measure of the size of the vector $x$, where eq$^n$ ① requires the size to be +ve, Eq$^n$ ② requires the size to be scaled as the vector is scaled, and Eq$^n$ ③ is known as the triangular inequality and has its origins in the notion of distance in $\mathbb{R}^3$.

⇒ Any mapping of an n-D vector space onto a subset of $\mathbb{R}$ that satisfies these 3 requirements can be called a "norm". The space together with a defined norm is called a Normed Linear Space.

Ex:- For the vector space $V = \mathbb{R}^n$ with $x \in V$ given by $x = (x_1, x_2, \dots, x_n)$, an obvious definition of a norm is :

$$\|x\| = \left(x_1^2 + x_2^2 + \dots + x_n^2\right)^{1/2}$$

Ex:- Another norm on $V = \mathbb{R}^n$ is

$$\|x\| = \max_{1 \le i \le n} \left\{ |x_i| \right\}$$

all 3 axioms are obeyed.

Normed Linear Spaces Examples

Vector Norm

- The linear Space $\mathbb{R}^n$ (Euclidean Space) where $x = (x_1, x_2, \dots, x_i, \dots, x_n)$ together with the norm:

$$\|x\|_p = \left(\sum_{i=1}^{p} |x_i|^p\right)^{1/p} \quad , p \geq 1$$

is known as the $L_p$ normed linear space.
The most common are the one norm, $L_1$ and the two norm, $L_2$, linear spaces where $p=1$ and $p=2$, respectively.

- The other standard norm for the space $\mathbb{R}^n$ is the infinity, or maximum norm given by

$$\|x\|_\infty = \max_{1 \leq i \leq n} (|x_i|)$$

The vector space $\mathbb{R}^n$ together with the infinity norm is commonly denoted $L_\infty$.

Ex: $x = (3, -1, 2, 0, 4) \in \mathbb{R}^5$

(i) One norm: $\|x\|_1 = \sum |x_i|$

$= |3| + |-1| + |2| + |0| + |4| = 10$

(ii) Two norm: $\|x\|_2 = \sqrt{|3|^2 + |-1|^2 + |2|^2 + |0|^2 + |4|^2} = \sqrt{30}$

(iii) Infinity norm: $\|x\|_\infty = \max(|3|, |-1|, |2|, |0|, |4|) = 4$

Note: Each is a different way of measuring ~~of~~ the size of a vector $\in \mathbb{R}^n$.

## Sub-ordinate Matrix Norm

* The norm of a matrix is a real # which is a measure of the magnitude of the matrix (how large its elements are).
It is a way of determining the "size" of a matrix that is not necessarily related to how many rows/columns the matrix has.

---

For the Normed Linear Space $\{\mathbb{R}^n, \|x\|\}$, where $\|x\|$ is some norm, we define the norm of the matrix $A_{n\times n}$ which is sub-ordinate to the vector norm $\|x\|$ as

$$\|A\| = \max_{x \neq 0} \left( \frac{\|Ax\|}{\|x\|} \right)$$

$x \in \mathbb{R}^n \Longrightarrow Ax \in \mathbb{R}^n$,

so $\|A\|$ is the largest value of the vector norm of $Ax$ normalized over all non-zero vectors $x$.

The 3 requirements of a vector norm are properties of $\|A\|$. There are 2 further properties which are a consequence of the definition for $\|A\|$. Hence, sub-ordinate matrix norms satisfy the following 5 rules:

$$\|A\| > 0 \quad , \quad A \neq 0,$$

$$\|\alpha A\| = |\alpha| \|A\| \quad , \quad \alpha \in \mathbb{R}$$

$$\|A + B\| \leq \|A\| + \|B\|$$

$$\|Ax\| \leq \|A\| \|x\|$$

$$\|AB\| \leq \|A\| \|B\|$$

The easiest matrix norms to compute are matrix norms sub-ordinate to the $L_1$ and $L_\infty$ vector norms. These are.

$$\|A\|_1 = \max_{x \neq 0} \left( \frac{\sum_{i=1}^{n} |(Ax)_i|}{\sum_{i=1}^{n} |x_i|} \right)$$

$$\|A\|_\infty = \max_{x \neq 0} \left( \frac{\max_{1 \le i \le n} |(Ax)_i|}{\max_{1 \le i \le n} (|x_i|)} \right)$$

Ex :

$$A = \begin{bmatrix} 3 & -6 & 2 \\ 2 & 5 & 1 \\ -3 & 2 & 2 \end{bmatrix}$$

$$\|A\|_1 = \max\left((|3|+|2|+|-3|), (|-6|+|5|+|2|), (|2|+|1|+|2|)\right)$$

$$= \max(8, 13, 5) = 13.$$

$$\|A\|_\infty = \max\left((|3|+|-6|+|2|), (|2|+|5|+|1|), (|-3|+|2|+|2|)\right)$$

$$= \max(11, 8, 7) = 11$$

Ex: 1. A rank-one matrix $A = uv^T$ is as basic as we can get. It has 1 non-zero eigenvalue $\lambda_1$ and 1 non-zero singular value $\sigma_1$.

Its eigenvector is $u$ and its singular vectors are $u$ and $v$.

Eigenvector: $Au = (uv^T)u = u(v^Tu) = (v^Tu)\,u = \lambda_1 u$

$$\text{So } \lambda_1 = v^Tu$$

Singular vector: $A^TAv = (vu^T)(uv^T)v = v(u^Tu)\,v^Tv$

$$= v(u^Tu)(v^Tv) = (u^Tu)(v^Tv)v = \sigma_1^2 v$$

$$\text{So } \sigma_1 = \|u\|\|v\|$$

$$|\lambda_1| \le \sigma_1 \Longleftrightarrow |v^Tu| \le \|u\|\|v\|$$

Schwarz inequality

$|A| = |\lambda_1 \lambda_2| = \sigma_1 \sigma_2$

$$|\lambda_1| \le \sigma_1 \implies \frac{|A|}{|\lambda_1|} \ge \frac{|A|}{\sigma_1}$$

$$\therefore |\lambda_2| \ge \sigma_2$$

## Low-rank approximation of matrices

Problem: For any matrix $A \in M_{m\times n}$ and integer $k \geq 1$, find the rank-$k$ matrix $B$ that is the closest to $A$?

$$\min_{\substack{B \in M_{m\times n} \\ \text{rank}(B) \leq k}} \|A - B\| = ?$$

### Eckart-Young-Mirsky theorem

Suppose $A \in M_{m\times n}$ has singular value decomposition (SVD), $A = U\Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T$.

and define $A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$, be the truncated SVD of $A$ with $k \leq r$. Then

$A_k$ is the best rank-$k$ approximation to $A$

$$\min_{\substack{B \in M_{m\times n} \\ r_B \leq k}} \|A - B\| = \|A - A_k\| = \sigma_{k+1}$$

$$\|A - B\| \geq \|A - A_k\| = \sigma_{k+1}$$

for all matrices $B$ of rank $k$.

Proof

Let $D_k$ be the diagonal matrix

$$D_k = \begin{bmatrix} 0 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & & & \\ & & & \sigma_{k+1} & & & \\ & & & & \ddots & & \\ & & & & & \sigma_r & \\ & & & & & & 0 \\ & & & & & & & 0 \end{bmatrix}$$

$$\|A - A_k\| = \left\| \sum_{i=1}^{r} \sigma_i u_i v_i^T - \sum_{i=1}^{k} \sigma_i u_i v_i^T \right\|$$

$$= \left\| \sum_{i=k+1}^{r} \sigma_i u_i v_i^T \right\|$$

$$= \|U D_k V^T\|$$

$$= \|D_k\| \qquad \left[ \text{since } U \text{ \& } V \text{ are orthogonal} \quad \|UD_kV^T\| = \max \frac{\|UD_kV^T x\|}{\|x\|} \right]$$

$$= \sigma_{k+1} \qquad = \max \frac{\|D_k x\|}{\|x\|} = \|D_k\|$$

Let $V_{k+1} = \begin{bmatrix} v_1 & \cdots & \cdots & v_{k+1} \end{bmatrix}$, where

$v_1, \ldots, v_{k+1}$ are the eigenvectors associated with the top $k+1$ singular values.

$v_i \in C(B)$

$B^T B v_i = \sigma_i^2 v_i$

$$\dim(\mathcal{N}(B)) + \mathrm{rank}(B) = n$$

$$\dim(\mathcal{N}(B)) + \dim(\mathcal{R}(V_{k+1})) > n$$

$$\Rightarrow \mathcal{N}(B) \cap \mathcal{R}(V_{k+1}) \neq \{0\}$$

We can take a unit vector $x \in N(B) \cap R(V_{k+1})$.

$$\|A-B\|^2 = \|A-B\|^2 \|x\|^2 \geq \|(A-B)x\|^2$$

$$= \|Ax\|^2 \quad \left[ \text{since } x \in N(B), \; Bx = 0 \right.$$

$$= \|UDV^T x\|^2$$

$$= \|DV^T x\|^2 \quad \left[ \text{since } U \text{ is orthogonal} \right.$$

$$= \left\| \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \begin{bmatrix} v_1^T \\ \vdots \\ v_r^T \end{bmatrix} x \right\|^2$$

$$= \left\| \sum_{i=1}^{r} \sigma_i v_i^T x \right\|^2$$

$$= \sum_{i=1}^{r} \sigma_i^2 (v_i^T x)^2$$

$$= \sum_{i=1}^{k+1} \sigma_i^2 (v_i^T x)^2 \quad \left[ x \in R(V_{k+1}), \; x \perp v_i, \; \forall i > k+1 \right.$$

$$\geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} (v_i^T x)^2$$

$$= \cancel{\sigma_{k+1}^2 \|V^T x\|^2} = \sigma_{k+1}^2 \sum_{i=1}^{r} (v_i^T x)^2$$

$$\cancel{\leq \sigma_{k+1}^2 \|} \qquad = \sigma_{k+1}^2 \|V^T x\|^2$$

$$= \sigma_{k+1}^2 \|x\|^2 \qquad \text{[ since } V \text{ is orthogonal}$$

$$= \sigma_{k+1}^2$$

$$= \|A - A_k\|^2.$$

For any matrix $B \in M_{m\times n}$

$$\Longrightarrow \underline{\|A-B\| \geq \|A-A_k\|}$$

# Polar Decomposition, $A = QS$

Every complex # $x+iy$ has the polar form $re^{i\theta}$. A number $r \geq 0$ multiplies a # $e^{i\theta}$ on the unit circle.

$$x+iy = r\cos\theta + i r\sin\theta = r(\cos\theta + i\sin\theta) = re^{i\theta}$$

$\Rightarrow$ 1 by 1 matrices

$\Downarrow$

$e^{i\theta}$ is an orthogonal matrix

and $r \geq 0$ is a +ve semidefinite matrix (call it $S$)

The polar decomposition extends the same idea to $n \times n$ matrices: orthogonal times +ve semidefinite.

> Every real square matrix can be factored into
> $A = QS$ = (orthogonal) $\times$ (symmetric positive semidefinite)

* If '$A$' is invertible, $S$ is +ve definite.

Proof

Insert $V^T V = I$ into the middle of SVD

$$\boxed{\text{Polar decomposition}: \quad A = U\Sigma V^T = (UV^T)(V\Sigma V^T) = QS}$$

$(UV^T)(UV^T)^T = UV^T V U^T = I \Rightarrow UV^T$ is orthogonal.

since $U$ & $V$ are orthogonal ($U^T U = I$ & $V^T V = I$)

The eigenvalues of $V\Sigma V^T$ are in $\Sigma$ $\therefore V\Sigma V^T$ is +ve semidefinite.

$\sigma_i \geq 0$

$A^T A v_i = \sigma_i^2 v_i$ & $AA^T u_i = \sigma_i^2 u_i$

* $S = V\Sigma V^T$ is the symmetric +ve definite square root of $A^T A$.

$S^2 = V\Sigma^2 V^T = A^T A$

$\Rightarrow$ Eigenvalues of $S$ are the singular values of $A$.
Eigenvectors of $S$ are the singular vectors $v$ of $A$.

* $Q = UV^T$ is the nearest orthogonal matrix to $A$.

$$\min_{Q^TQ=I} \|A-Q\| = \|A-UV^T\|$$

$$\|A-Q\| \geq \|A-UV^T\| \quad \text{for all orthogonal matrices } Q.$$

$$A_{m\times n} = U_{m\times m} \Sigma_{m\times n} V_{n\times n} \implies (m=n)$$

- $e^{i\theta}$ is the nearest number on the unit circle to $re^{i\theta}$.

**Proof**

$$\|A-Q\| = \|U\Sigma V^T - Q\| = \|U^T(U\Sigma V^T - Q)V\|$$

$$= \|\Sigma - U^TQV\| = \|\Sigma - Q'\|$$

$$(Q')^TQ' = (U^TQV)^T(U^TQV) = V^TQ^TUU^TQV = V^TQ^TQV = I.$$

$Q'$ is orthogonal.

We want to minimize $\|\Sigma - Q'\|$ over all orthogonal matrices $Q'$.

$$|a-b| \geq \big||a|-|b|\big|$$

reverse triangle inequality

$$\left.\begin{aligned} |a| &= |a-b+b| \leq |a-b|+|b| \implies |a|-|b| \leq |a-b| \\ |b| &= |b-a+a| \leq |a-b|+|a| \implies |a|-|b| \geq -|a-b| \end{aligned}\right\} \implies -|a-b| \leq |a|-|b| \leq |a-b|$$

$$\cancel{\|\Sigma - Q'\| = \max}$$

$$\|A\| = \max_{|x| \neq 0} \frac{\|Ax\|}{\|x\|} = \sigma_1$$

For

$$\max \frac{\|A(tx)\|}{\|tx\|} = \max \frac{t\,\|Ax\|}{t\,\|x\|} = \max \frac{\|Ax\|}{\|x\|}$$

$\Longrightarrow$ scaling invariant.

$$\|\Sigma - Q'\| = \max_{\|x\|=1} \frac{\|(\Sigma - Q')x\|}{\cancel{\|x\|}}$$

$$= \max_{\|x\|=1} \|\Sigma x - Q'x\|$$

$$\longrightarrow \geq \max_{\|x\|=1} \big|\|\Sigma x\| - \|Q'x\|\big|$$

$$\longrightarrow = \max_{\|x\|=1} \big|\|\Sigma x\| - \|x\|\big|$$

$$= \max_{\|x\|=1} \big|\|\Sigma x\| - 1\big|$$

$$= |\sigma_1 - 1| = \|\Sigma - I\|$$

$$\Rightarrow \|A - Q\| = \|\Sigma - Q'\| = \|\Sigma - U^T Q V\| \geq \|\Sigma - I\|$$

$$\|U(\Sigma - I)V^T\| = \|A - UV^T\|$$

$$\therefore \|A - Q\| \geq \|A - UV^T\|$$

$\Longrightarrow$ $Q = UV^T$ is the nearest orthogonal matrix to $A = U\Sigma V^T$.

* If $A_{n\times n}$ is singular,

then the distance to a closest singular matrix is the smallest singular value $\sigma_n$.

ie.,

$\sigma_n$ is measuring the distance from A to singularity.

Proof

$$\min_{\substack{|B| \neq 0 \\ \text{rank}(B) = k < r}} \|A - B\| = \|A - A_k\| = \sigma_{k+1}$$

which is smallest when $k+1 = n$

$\therefore$

Ex: 2 $A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} = U \Sigma V^T$

Find polar decomposition $A = QS$

Ans: $Q = UV^T = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}$

$S = V \Sigma V^T = \sqrt{5} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

$A = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T$

$\begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} = 3\sqrt{5} \cdot \frac{1}{\sqrt{20}} \begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix} + \sqrt{5} \cdot \frac{1}{\sqrt{20}} \begin{bmatrix} 3 & -3 \\ -1 & 1 \end{bmatrix}$

$\sigma_2 = \sigma_{min} = \sqrt{5}$

Change $\sigma_{min} = \sigma_2 = 0$,

this knocks out the smallest piece in $A$. Then only the rank-1 (singular) matrix $\sigma_1 u_1 v_1^T$ will be left: the closest to $A$.

## The Condition Number of A

Some systems are sensitive, others are not so sensitive to roundoff errors.
The sensitivity to error is measured by the condition number.

$$Ax = b$$

Suppose the right side is changed to $b + \Delta b$ because of roundoff or measurement error. The solution is then changed to $x + \Delta x$.

Our goal is to estimate the change $\Delta x$ in the solution from the change $\Delta b$ in the equation.

$$A(x+\Delta x) = b + \Delta b$$

$$A x = b$$

---

$$A(\Delta x) = \Delta b \quad : \text{ Error equation.}$$

The error is, $\Delta x = A^{-1}(\Delta b)$

It is large when $A^{-1}$ is large.
(then $A$ is nearly singular).

$A^{-1}$ is large $\Longrightarrow \frac{1}{\sigma_1}$ is large $\Longrightarrow \sigma_1$ is small

$$\det(A^T A) = 0 = \det(A)\det(A^T) = [\det(A)]^2$$

$\Longrightarrow$ A is singular. (nearly)

: $\div$ :

$$\|Ax\| \leq \|A\| \, \|x\|$$

$$\Delta x = A^{-1}(\Delta b)$$

$$\|\Delta x\| = \|A^{-1}(\Delta b)\| \leq \|A^{-1}\| \, \|\Delta b\|$$

The worst error has, $\|\Delta x\| = \|A^{-1}\| \cdot \|\Delta b\|$

$\cancel{\Delta x}$ .

The error bound $\|A^{-1}\|$ has one serious drawback. If we multiply 'A' by 1000, then $A^{-1}$ is divided by 1000.

The matrix looks a 1000 times better. But a simple ~~[illegible]~~ rescaling cannot change the reality of the problem.

It is true that $\Delta x$ will be divided by 1000, but so will the exact solution $x = A^{-1}b$. The relative error $\frac{\|\Delta x\|}{\|x\|}$ will stay the same. It is this relative ~~error~~ change in '$x$' that should be compared to the relative change in '$b$'.

Comparing relative errors will now lead to the condition number, $\boxed{c = \|A\|\,\|A^{-1}\|}$.

Xing '$A$' by 1000 does not change this #, ~~than~~ The condition # $c$ measures the sensitivity of $Ax = b$.

* $c \geq 1$

PS. 11.2(4)

$$\left.\begin{aligned} \Delta x &= A^{-1}(\Delta b) \\ b &= A x \end{aligned}\right\}$$

$$\|\Delta x\| = \|A^{-1}(\Delta b)\| \le \|A^{-1}\| \cdot \|\Delta b\|$$

$$\|b\| = \|A x\| \le \|A\| \, \|x\|$$

$$\|\Delta x\| \, \|b\| \le \|A\| \, \|A^{-1}\| \, \|x\| \, \|\Delta b\|$$

$$\frac{\|\Delta x\|}{\|x\|} \le \|A\| \, \|A^{-1}\| \frac{\|\Delta b\|}{\|b\|}$$

The solution error is less than $c = \|A\| \|A^{-1}\|$ times the problem error:

$$\boxed{\begin{aligned} &\frac{\|\Delta x\|}{\|x\|} \le c \frac{\|\Delta b\|}{\|b\|} \\ &c = \|A\| \|A^{-1}\| \text{ is the condition number.} \end{aligned}}$$

$$(A+\Delta A)(x+\Delta x) = b$$

$$Ax = b$$

---

$$\Delta A(x+\Delta x) + A\Delta x = 0$$

$$A(\Delta x) = -(\Delta A)(x+\Delta x)$$

$$\Delta x = -A^{-1}(\Delta A)(x+\Delta x)$$

$$\|\Delta x\| \le \|A^{-1}\| \|\Delta A\| \|x+\Delta x\|$$

$$\frac{\|\Delta x\|}{\|x+\Delta x\|} \le \|A^{-1}\| \|\Delta A\| = \|A\| \|A^{-1}\| \frac{\|\Delta A\|}{\|A\|}$$

If the problem error is $\Delta A$ (error in $A$ instead of $b$), still $c$ controls $\Delta x$:

$$\boxed{\frac{\|\Delta x\|}{\|x+\Delta x\|} \le c \frac{\|\Delta A\|}{\|A\|}}$$

$A\| \|A^{-1}\|$

⇒ Errors enter in 2 ways. They begin with an error $\Delta A$ or $\Delta b$ — a wrong matrix or a wrong $b$. This problem error is amplified (a lot or a little) into the solution error $\Delta x$. That error is bounded, relative to $x$ itself by the condition number, $c$.

The error $\Delta b$ depends on computer round off and on the original measurements of $b$. The error $\Delta A$ also depends on the elimination steps.

When $\Delta A$ or the condition # is very large, the error $\Delta x$ can be unacceptable.

Ex: 3 When 'A' is symmetric, $c = \|A\| \|A^{-1}\|$ comes from the eigenvalues:

> * symmetric matrices are always diagonalizable
> * Eigenvectors of real symmetric matrix (corr. to diff. eigenvalues) are always $\perp$

$$A = \begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix}$$ has norm 6.

$$A^{-1} = \begin{bmatrix} 1/6 & 0 \\ 0 & 1/2 \end{bmatrix}$$ has norm $1/2$

This $A$ is symmetric +ve definite.

Its norm is $\lambda_{max} = 6$.

The norm of $A^{-1}$ is $\dfrac{1}{\lambda_{max}} = \dfrac{1}{2}$.

The condition number, $\|A\| \|A^{-1}\| = \dfrac{\lambda_{max}}{\lambda_{min}}$

Condition number for positive definite $A$:

$$\boxed{c = \|A\| \|A^{-1}\| = \frac{\lambda_{max}}{\lambda_{min}}}$$

* Symmetric matrices that have +ve eigenvalues are called Positive definite.

Ex:4. Keep the same $A$, with eigenvalues 6 & 2.

To make $x$ small, choose $b$ along the 1st eigenvector $(1,0)$.

Then,

$x = \frac{1}{6} b$ and $\Delta x = \frac{1}{2} \Delta b$

$\frac{\|\Delta x\|}{\|x\|} = 3$, is exactly $c = 3$ times the ratio $\frac{\|\Delta b\|}{\|b\|}$.

P.S (11.2)

1. Find the norms $\|A\| = \lambda_{max}$ & condition numbers $c = \frac{\lambda_{max}}{\lambda_{min}}$ of these positive definite matrices:

(a) $\begin{bmatrix} 0.5 & 0 \\ 0 & 2 \end{bmatrix}$

Ans: $\|A\| = 2$ & $\|A^{-1}\| = 2$

$$c = \|A\| \|A^{-1}\| = \frac{\lambda_{max}}{\lambda_{min}} = 4$$

(b) $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

Ans: $\|B\| = \lambda_{max} = 3$, $\|B^{-1}\| = \frac{1}{\lambda_{min}} = \frac{1}{1} = 1$

$c = 3$

2. Find the norms & condition numbers from the square roots of $\lambda_{max}(A^T A)$ and $\lambda_{min}(A^T A)$. Without the definiteness in $A$, we go to $A^T A$

(a) $\begin{bmatrix} -2 & 0 \\ 0 & 2 \end{bmatrix}$

Ans: $A^T A = \begin{bmatrix} -2 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} -2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$

$\|A\| = 2$ & $\|A^{-1}\| = \frac{1}{2}$

$C = \|A\| \|A^{-1}\| = 2 \times \frac{1}{2} = 1$

(b) $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$

Ans: $B^T B = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

$\det [B^T B] = 0$

$C = \infty$

(c) $\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$

$C^T C = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$

$\|A\| = \sqrt{2}$, $\|A^{-1}\| = \frac{1}{\sqrt{2}}$

$C = 1$

4. Use $\|AA^{-1}\| \le \|A\|\|A^{-1}\|$ to prove that the condition # is at least 1.

Ans: $\|AA^{-1}\| = \|I\| = 1 \le \|A\|\|A^{-1}\| = C$

$$\underline{C \geq 1}$$

5. Why is $I$ the only symmetric +ve definite matrix that has $\lambda_{max} = \lambda_{min} = 1$? Then the only other matrices with $\|A\| = 1$ and $\|A^{-1}\| = 1$ must have $A^T A = I$. Those are ________ matrices: perfectly conditioned.

Ans:

$$\lambda_{max} = 1 = \lambda_{min} \Longrightarrow \text{all } \lambda_i = 1$$

$$A = S I S^{-1} = \underline{I}$$

Only other matrices with $\|A\| = \|Q^{-1}\| = 1$ are orthogonal matrices.

6. Orthogonal matrices have norm $\|Q\|=1$. If $A=QR$ show that $\|A\|\le\|R\|$ and also $\|R\|\le\|A\|$. Then $\|A\|\le\|Q\|\|R\|$. Find an example of $A=LU$ with $\|A\|<\|L\|\|U\|$

Ans: $A=QR$

$$\|A\|=\|QR\|\le\|Q\|\cdot\|R\|=\|R\|$$

$$R=Q^{-1}A$$

$$\|R\|=\|Q^{-1}A\|\le\|Q^{-1}\|\,\|A\|=\|A\|$$

$$\|A\|\le\|R\| \quad \& \quad \|R\|\le\|A\|$$

$$\Longrightarrow \underline{\underline{\|A\|=\|R\|}}$$

7. (a) Which famous inequality gives,
$\|(A+B)x\| \le \|Ax\| + \|Bx\|$ for every $x$?
(b) Why does the definition of matrix norms lead to $\|A+B\| \le \|A\| + \|B\|$?

Ans: Triangle inequality

$$\max_{x \neq 0} \frac{\|(A+B)x\|}{\|x\|} \le \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} + \max_{x \neq 0} \frac{\|Bx\|}{\|x\|}$$

$$\|A+B\| \le \|A\| + \|B\|$$

8. Show that if $\lambda$ is an eigenvalue of $A$, then $|\lambda| \le \|A\|$. Start from $Ax = \lambda x$

Ans: $|\lambda| \|x\| = \|Ax\| \le \|A\| \|x\|$

$$\|A\| \ge |\lambda|$$

9. The "spectral radius" $\rho(A) = |\lambda_{max}|$ is the largest absolute value of the eigenvalues. Show with 2×2 examples that $\rho(A+B) \le \rho(A)+\rho(B)$ and $\rho(AB) \le \rho(A)\,\rho(B)$ can both be false. The spectral radius is not acceptable as a norm

Ans:

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

$$\left.\begin{array}{l} \rho(A) = 0,\ \rho(B) = 0 \\ \rho(A+B) = 1 \end{array}\right\} \quad \underline{\underline{\rho(A+B) = 1 < 0+0 = \rho(A) + \rho(B)}}$$

The triangle inequality $\|A+B\| \le \|A\| + \|B\|$ fails for the spectral radius

$$AB = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\rho(AB) = 1 > 0 = \rho(A)\,\rho(B)$$

$\Rightarrow$ $\rho(A) = |\lambda_{max}|$ = spectral radius is not a norm.

11. Estimate the condition # of the ill conditioned matrix $A = \begin{bmatrix} 1 & 1 \\ 1 & 1.0001 \end{bmatrix}$

$A$ is +ve definite

Ans: $C(A) = \dfrac{\lambda_{max}}{\lambda_{min}}$

$$= \frac{20001 + \sqrt{400000001}}{20000} \Big/ \frac{20001 - \sqrt{400000001}}{20000}$$

$$= \frac{20001 + 20000.000025}{20001 - 20000.000025}$$

$$\approx \underline{\underline{40,000}}$$

12. Why is the determinant of 'A' no good as a norm? Why is it no good as a condition #?

Ans: $\det(\alpha A) \neq \alpha \det(A)$

$\det(A+B)$ is not always less than $\det(A) + \det(B)$

$\det(AB) = \det(A)\det(B)$ is the only reasonable property.

The condition # should not change when $A$ is multiplied by 10.

13.

15. The "$l^1$ norm" and the "$l^\infty$ norm" of $x = (x_1, \dots, x_n)$ are:

$$\boxed{\|x\|_1 = |x_1| + \dots + |x_n|}$$

$$\boxed{\|x\|_\infty = \max_{1 \le i \le n} |x_i|}$$

Compute the norms $\|x\|$, $\|x\|_1$, $\|x\|_\infty$ of these 2 vectors in $\mathbb{R}^5$.

(a) $x = (1, 1, 1, 1, 1)$,

Ans: $\|x\| = \sqrt{5}$

$\|x\|_1 = 5$

$\|x\|_\infty = 1$

(b) $x = (0.1, 0.7, 0.3, 0.4, 0.5)$

Ans: $\|x\| = 1$

$\|x\|_1 = 2$

$\|x\|_\infty = 0.7$

16. ~~Sketch~~

Prove that $\|x\|_\infty \le \|x\| \le \|x\|_1$.

Show from the Schwarz inequality that the ratios $\frac{\|x\|}{\|x\|_\infty}$ and $\frac{\|x\|_1}{\|x\|}$ are never larger than $\sqrt{n}$. Which vector $(x_1, \dots, x_n)$ gives ratios equal to $\sqrt{n}$?

Ans: $x_1^2 + \dots + x_n^2 \ge \max(x_i^2)$

$x_1^2 + \dots + x_n^2 \le (|x_1| + \dots + |x_n|)^2$

$\|x\|_\infty \le \|x\| \le \|x\|_1$

---

$x_1^2 + \dots + x_n^2 \le n \cdot \max(x_i^2)$

$$\boxed{\|x\| \le \sqrt{n}\, \|x\|_\infty}$$

Schwarz inequality,

$$x \cdot y \le \|x\|\|y\|$$

~~Set $y = (1, 1, \cdots, 1)$~~

Set $y_i = \operatorname{sgn}(x_i)$.

$$\|x\|_1 = x \cdot y \le \|x\|\|y\| = \sqrt{n}\,\|x\|$$

$$\boxed{\|x\|_1 \le \sqrt{n}\,\|x\|}$$

17. All vector norms must satisfy the triangle inequality. Prove that

$$\|x+y\|_\infty \le \|x\|_\infty + \|y\|_\infty$$

$$\|x+y\|_1 \le \|x\|_1 + \|y\|_1$$

Ans: For the $l^{\infty}$ norm, the largest component of $x$ + the largest component of $y$ is not less than $\|x+y\|_{\infty}$ = the largest component of $x+y$.

For the $l^{1}$ norm, each component has

$$|x_i + x_j| \leq |x_i| + |x_j|$$

Sum $\Rightarrow \|x+y\|_1 \leq \|x\|_1 + \|y\|_1$

18. Vector norms must also satisfy $\|cx\| = |c|\,\|x\|$. The norm must be +ve except when $x=0$. Which of these are norms for vectors $(x_1, x_2)$ in $\mathbb{R}^2$?

(a) $\|x\|_A = |x_1| + 2|x_2|$

Ans: $\|x\|_A > 0$, $x \neq 0$

$$\|\alpha x\|_A = |\alpha x_1| + 2|\alpha x_2| = \alpha\left[|x_1| + 2|x_2|\right]$$

$$= \alpha \|x\|_A$$

$$\|x+y\| = |x_1+y_1| + 2|x_2+y_2| \le |x_1|+|y_1| + 2|x_1| + 2|y_1|$$

$$\cancel{\|x\| + \|y\| = |x_1| +} \cdot \le \|x_1\| + \|y_1\|$$

$\Longrightarrow$

$\|x\|_A = |x_1| + 2|x_2|$ is a norm

(b) $\|x\|_B = \min(|x_1|, |x_2|)$

Ans: $\|x\|_B \geq 0$, $x \neq 0$

$$\|\alpha x\|_B = \min(|\alpha x_1|, |\alpha x_2|)$$

$$= |\alpha| \min(|x_1|, |x_2|)$$

$$\|x+y\| = \min(|x_1+y_1|, |x_2+y_2|)$$

need not have to be less than

$\|x\| + \|y\| = \min(|x_1|, |y_1|) + \min(|x_2|, |y_2|)$

(c)

Ans:

$2|x_1| + 2|y_1|$

Ex:-

$\cancel{x = (+2, -3), \quad y = (+2, -4)}$

$\cancel{\|x+y\| = \min(4, 7) \neq 4}$

$\cancel{\|x\| + \|y\| = \min(}$

Ex:- $x = (4, -5)$, $y = (3, -1)$

$$\|x+y\| = \min(7, 6) = 6$$

$$\|x\| + \|y\| = 4 + 1 = 5$$

$$\|x+y\| > \|x\| + \|y\|$$

$\Longrightarrow \|x\|_B = \min(|x_1|, |x_2|)$ is not a norm.

(c) $\|x\|_c = \|x\| + \|x\|_\infty$

Ans: $\|x\|_c > 0, \quad x \neq 0$

$\|\alpha x\|_c = \alpha \|x\|_c$

$\left.\begin{matrix} \\ \\ \end{matrix}\right\}$ $\|x\|_c$ is a norm

(d) $\|x\|_D = \|Ax\|$

Ans: $\|x\|_D = \|Ax\| > 0$ , $x \neq 0$

given 'A' is invertible.

~~Ax+y~~

$$\|\alpha x\|_D = \|A\alpha x\| = \alpha\|Ax\|$$

$$\|x+y\|_D = \|A(x+y)\| = \|Ax + Ay\| \le \|Ax\| + \|Ay\|$$

$$\le \|x\|_D + \|y\|_D$$

19. Show that $x^T y \le \|x\|_1 \|y\|_\infty$ by choosing components $y_i = \pm 1$ to make $x^T y$ as large as possible.

Ans: $x^T y = x_1 y_1 + x_2 y_2 + \cdots \le (\max|y_i|)(|x_1| + |x_2| + \cdots)$

$$= \|y\|_\infty \|x\|_1$$

# The Pseudoinverse $A^+$

By choosing good bases,

$A$ multiplies $v_i$ in the row space to give $\sigma_i u_i$ in the column space. $A^{-1}$ must do the opposite — If $Av = \sigma u$ then $A^{-1}u = \frac{v}{\sigma}$.
The singular values of $A^{-1}$ are $\frac{1}{\sigma}$, just as the eigenvalues of $A^{-1}$ are $1/\lambda$. The bases are reversed. The $u$'s are in the row space of $A^{-1}$, the $v$'s are in the column space.

If $A^{-1}$ exists !

A matrix that multiplies $u_i$ to produce $\frac{v_i}{\sigma_i}$ does exist. It is the pseudoinverse $A^+$.

Pseudoinverse of A,

$$A^{+} = V\Sigma^{+}U^{T} = \underbrace{\begin{bmatrix} v_1 \cdots v_r \cdots v_n \end{bmatrix}}_{(n \times n)} \underbrace{\begin{bmatrix} \sigma_1^{-1} & & & \\ & \ddots & & \\ & & \sigma_r^{-1} & \\ & & & \end{bmatrix}}_{(n \times m)} \underbrace{\begin{bmatrix} u_1 \cdots u_r \cdots u_m \end{bmatrix}^{T}}_{(m \times m)}$$

$$= \frac{1}{\sigma_1} v_1 u_1^T + \frac{1}{\sigma_2} v_2 u_2^T + \cdots\cdots + \frac{1}{\sigma_r} v_r u_r^T$$

$$AA^T u_i = \sigma_i^2 u_i$$
$$A^T A v_i = \sigma_i^2 v_i$$

$$Av_i = \sigma_i u_i$$

If $A^{-1}$ exists, then $A^{+}$ is the same as $A^{-1}$.
In that case $m = n = r$.

When $r < m$ & $r < n$, $A^{+}$ is needed.
Then 'A' has no 2-sided inverse, but it has a pseudoinverse $A^{+}$ with that same rank $r$:

$$\boxed{\begin{aligned} &A^{+}u_i = \frac{1}{\sigma_i} v_i \quad \text{for } i \le r \text{ and} \\ &A^{+}u_i = 0 \quad \text{for } i > r \end{aligned}}$$

- The vectors $u_1, \dots, u_r \in C(A)$ go back to $v_1, \dots, v_r \in C(A^T)$.

The other vectors $u_{r+1}, \dots, u_m \in N(A^T)$, and $A^+$ sends them to zero.

- Each $\sigma$ in $\Sigma$ is replaced by $\sigma^{-1}$ in $\Sigma^+$.

The product $\Sigma^+ \Sigma$ is as near to the identity as we can get. It is a projection matrix.

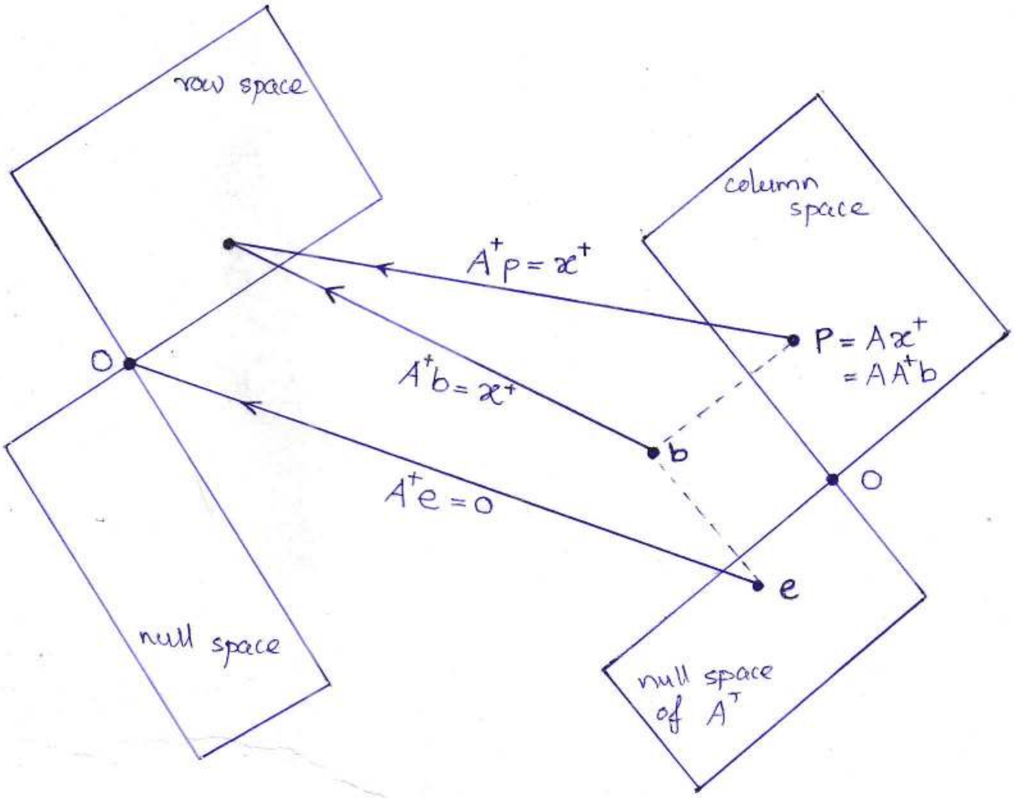
$\Sigma^+ \Sigma$ is partly $I$ & otherwise zero.

Ex:- $\sigma_1 = 2, \ \sigma_2 = 3$

$$\Sigma^+ \Sigma = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$$

$A$ : Rowspace to column space

$A^+$ : Columnspace to row space.




Pseudoinverse, $A^+$

* $Ax^+$ in the column space goes back to $A^+Ax^+ = x^+$ in the row space.

* Pseudoinverse $A^+$ is the unique matrix satisfying the Moore-Penrose conditions:

$$AA^+A = A$$

$$A^+AA^+ = A^+$$

$$(AA^+)^T = AA^+$$

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

Proof

1) $AA^+A = U\Sigma \underbrace{V^T V}_{I} \Sigma^+ \underbrace{U^T U}_{I} \Sigma V^T = U\Sigma\Sigma^+\Sigma V^T$

$= U\Sigma V^T = A$

3) $(AA^+)^T = (U\Sigma V^T V \Sigma^+ U^T)^T = (U\Sigma\Sigma^+U^T)^T = U(\Sigma\Sigma^+)^T U^T$

$= U(\Sigma\Sigma^+)U^T = U\Sigma V^T V\Sigma^+U^T = AA^+$

Identities

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

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

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

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

$$A^T = A^T A A^+$$

$$A^T = A^+ A A^T$$

Proof.

① $A^+ = A^+ A A^+$ & $(A A^+)^T = A A^+$

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

③ $A = A A^+ A$ and $A A^+ = (A A^+)^T$

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

(5) $A = AA^+A$ and $A^+A = (A^+A)^T$

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

* If $\operatorname{rank}(A) = n$,

$$A^+ = (A^T A)^{-1} A^T \quad \text{and} \quad A^+ A = I_n$$

If $\operatorname{rank}(A) = m$,

$$A^+ = A^T (A A^T)^{-1} \quad \text{and} \quad A^* A^+ = I_m$$

Proof

① $\operatorname{rank}(A_{m\times n}) = ~~\text{rank}~~\ n = \operatorname{rank}\left((A^T A)_{n\times n}\right)$

$\therefore$ $A^T A$ is invertible.

$$A^T = A^T A A^+ \implies \underline{\underline{A^+ = (A^T A)^{-1} A^T}}$$

② $\operatorname{rank}(A_{m\times n}) = m = \operatorname{rank}\left[(A A^T)_{m\times m}\right]$

$\therefore$ $A A^T$ is invertible

$$A^T = A^+ A A^T \implies \underline{\underline{A^+ = A^T (A A^T)^{-1}}}$$

a1
7.4

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T \qquad AA^+ = \sum_{i=1}^{r} u_i u_i^T$$

$$A^+ = \sum_{i=1}^{r} \frac{v_i u_i^T}{\sigma_i} \qquad A^+A = \sum_{i=1}^{r} v_i v_i^T$$

*

Starts 31/10/2020

$P_{C(A)} = AA^+$ : projection matrix onto $C(A)$

$$\Longrightarrow P_{N(A^T)} = I_m - AA^+$$

$P_{C(A^T)} = A^+A$ : projection matrix onto $C(A^T)$

$$\Longrightarrow P_{N(A)} = I_n - A^+A$$

**Proof**

A linear transformation $P$ is called an orthogonal projection if the image of $P$ is $V$ and the kernel is $\perp$ to $V$ and $P^2 = P$.

ILA(5)

$$\boxed{P^T = P \quad \& \quad P^2 = P}$$

If '$P$' is a projection matrix for an orthogonal projection, then for all $x, y \in \mathbb{R}^n$

$$Px \perp y - Py$$

$$0 = (Px)^T(y - Py) = x^T P^T(y - Py) = x^T P^T(I - P)y$$

$$0 = x^T(P^T - P^T P)y \quad \text{for all } x, y \in \mathbb{R}^n$$

$\therefore \quad P^T = P^T P$, hence

$$P = (P^T)^T = (P^T P)^T = P^T P = P^T \Longrightarrow \underline{P^T = P}$$

- $P^2 = (AA^+)^2 = AA^+AA^+ = U\Sigma V^T V \Sigma^+ U^T U \Sigma V^T V \Sigma^+ U^T$

  $= U\Sigma\Sigma^+\Sigma\Sigma^+U^T =$

  $\cancel{A^+AA^+ = V}$

- $AA^+A = A$

  $P^2 = (AA^+)^2 = \underbrace{AA^+A}A^+ = AA^+ = P$

  $\longrightarrow$ idempotent.

- $(AA^+)^T = AA^+$

$\Rightarrow$ $P = AA^+$ is an orthogonal projection

- If $y \in C(A)$, i.e., $y = Ax$ for some $x$

  $PA = AA^+A = A$

  $AA^+y = \underline{Py} = PAx = Ax = \underline{y} \Rightarrow y \in C(AA^+)$

  $\therefore$ $C(A) \subset C(AA^+)$

Conversely,

if $Py = AA^+y = y$, ie., $y \in C(AA^+)$

$y = A(A^+y) \Rightarrow y \in C(A)$

$\therefore C(AA^+) \subset C(A)$

$\Rightarrow \underline{C(AA^+) = C(A)}$

$\Longrightarrow P = AA^+$ is the orthogonal projector onto the $C(A)$.

$I - AA^+$ is the orthogonal projector onto the orthogonal complement of $C(A)$, ie., $N(A^T)$.

* $C(A^+) = C(A^T)$

Stud: 1/11/20

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

Proof $A = U\Sigma V^T$

$$\left.\begin{aligned} A^+ &= V\Sigma^+U^T \\ A^T &= V\Sigma^T U^T \end{aligned}\right\}$$

[OR]

Ex:3 Every rank-1 matrix is a column times a row. With unit vectors u and v,

$A = \sigma u v^T$. Its pseudo inverse is $A^+ = \frac{v u^T}{\sigma}$

$A A^+ = u u^T$, the projection onto the line thro' u.

$A^T A = v v^T$

Ex:4. Find the pseudo-inverse of $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.

Ans: $|A| = 0 \Rightarrow A$ is not invertible

rank$(A) = 1$

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

$\lambda_1 \lambda_2 = 0$ & $\lambda_1 + \lambda_2 = 4 \Rightarrow$ ~~$\sigma = 2$~~

~~$\sigma_1^2 = 2$~~ $\lambda = 0, 4 \Rightarrow \sigma_1^2 = 0, 4$

$v_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$A^+ = V \Sigma^+ U^T = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1/2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \frac{1}{\sqrt{2}}$

$= \frac{1}{2} \begin{bmatrix} 1/2 & 0 \\ 1/2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \Rightarrow$ rank$(A^+) = 1$

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

## Least Squares with Dependent Columns

check ILA (5)

Which $\hat{x}$ is the best if 'A' has dependent columns ?

$L: x_1 + x_2 t = b$

points : $(1,3)$, $(1,1)$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} = b$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$A\hat{x} = p$$

$$p = Pb = \frac{aa^T}{a^Ta} b$$

$$= \frac{1}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 4 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$Ax = b \implies A\hat{x} = p$$

Eq$^n$ with no solution — Eq$^n$ with infinitely many solutions.

The problem is that 'A' has dependent columns, and $e = b - p = (3,1) - (2,2) = (1,-1)$ is in its left ~~column~~ null space.

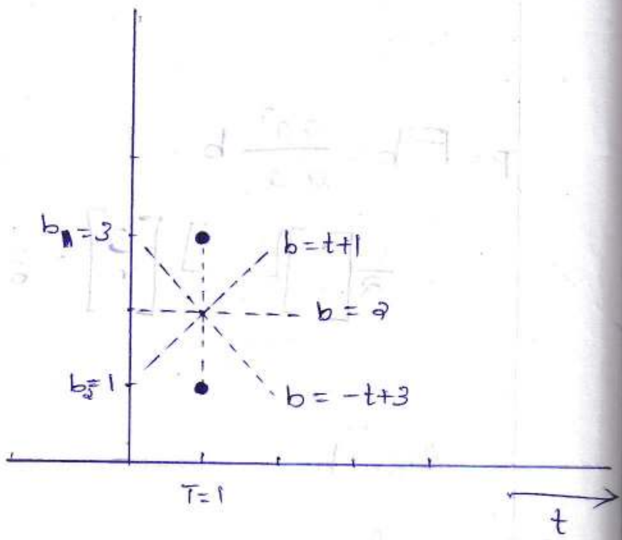
$$x_1 + x_2 = 2 \longrightarrow \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \hat{x}_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$(\hat{x}_1, \hat{x}_2) = \underset{b=2}{(2,0)},\ \underset{b=t+1}{(1,1)},\ \underset{b=-t+3}{(3,-1)},\ \cdots$$

All these lines have the same 2 errors

Those errors $e = b - p = (1,-1)$ are as small as possible. But this doesn't tell us which dash line is the best.

~~$1+t=b$~~
~~$1+t=3b$~~




The measurements $b_1 = 3$ and $b_2 = 1$ are at the same time $T=1$.

$\Rightarrow$ straight line $C+Dt$ can not go thro' both points.

$A^T A$ is singular

$$Ax = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} = b$$

unsolvable

$$A^T A \hat{x} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix} = A^T b$$

infinitely solvable

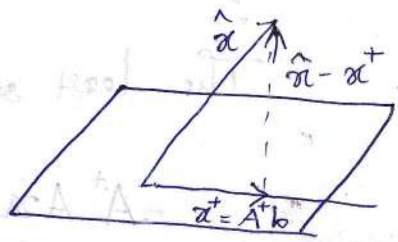
$\hat{x}_1 + \hat{x}_2 = 2.$

Any vector $\hat{x} = (1+c, 1-c)$ will solve those normal equations $A^T A \hat{x} = A^T b$.
The purpose of the pseudoinverse is to choose one solution $\hat{x} = x^+$.

$x^+ = A^+ b$ is the shortest solution to $A^T A \hat{x} = A^T b$ and $A\hat{x} = p$

$\hat{x} - x^+ \in N(A^T A) \Longrightarrow \hat{x} - x^+ \in N(A)$

$\hat{x} - x^+ \perp x^+$




$$\boxed{\|\hat{x}\|^2 = \|x^+\|^2 + \|\hat{x} - x^+\|^2} \Longrightarrow \boxed{\|\hat{x} - x^+\|^2 = \|\hat{x}\|^2 - \|x^+\|^2}$$

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

$$AA^T = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$$

Check Ex: 4

$$A^+ = \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$x^+ = A^+ b = \frac{1}{4}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ 1 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 4 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

---

$$Ax = b \implies Ax \in C(A)$$

chegg * The optimal $x$ should be such that $Ax = AA^+b$, since $P = AA^+$ is the projection operator onto the $C(A)$.

$\therefore$ The least square solution will be

$$A^+Ax = A^+AA^+b \implies x^+ = A^+b$$

$\hat{x} = (1+c, 1-c)$

$|\hat{x}|^2 = (1+c)^2 + (1-c)^2 = 2 + 2c^2$

is shortest when $c = 0$.

$\Longrightarrow$ $x^+ = (1,1) \in C(A^T)$ which is shorter than any other solution. $\hat{x} = (1+c, 1-c)$.

* The pseudoinverse $A^+$ and this best solution $x^+$ are essential in statistics, because experiments often have a matrix with dependent columns as well as dependent columns.

$\boxed{7.4 (A)}$ If 'A' has rank $n$ (full column rank) then it has a left inverse $L = (A^TA)^{-1}A^T$.

$LA = I$. Explain why the pseudoinverse is $A^+ = L$ in this case.

If 'A' has rank $m$ (full row rank) then it has a right inverse $R = A^T(AA^T)^{-1}$. This matrix $R$ gives $AR = I$. Explain why the pseudoinverse is $A^+ = R$ in this case.

Find $L$ for $A_1$ and $R$ for $A_2$.

Find $A^+$ for all 3 matrices $A_1, A_2, A_3$:

$A_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$, $A_2 = \begin{bmatrix} 2 & 2 \end{bmatrix}$, $A_3 = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}$

Ans: If 'A' has independent columns, then $A^TA$ is invertible.

$\Longrightarrow L = (A^TA)^{-1}A^T$ multiplies $A$ to give $LA = I$

$AL = A(A^TA)^{-1}A^T$ is the projection matrix on the column space.

$LA$ and $AL$ are projections on $C(A)$ & $C(A^T)$

If $A$ has rank $m$ (full row rank) then $AA^T$ is invertible.

$A$ multiplies $R = A^T(AA^T)^{-1}$ to give $AR = I$.

$RA = A^T(AA^T)^{-1}A$ is the projection matrix onto the row space, $C(A^T)$.

$\therefore \quad R = A^+$

$$A_1^+ = (A_1^T A_1)^{-1} A_1^T = \frac{1}{\sqrt{8}}\begin{bmatrix} 2 & 2 \end{bmatrix}$$

$$A_2^+ = A_2^T (A_2 A_2^T)^{-1} = \frac{1}{\sqrt{8}}\begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$A_1^+ A_1 = [1]$ and $A_2 A_2^+ = [1]$

But,

$A_3 = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}$ has no left or right inverse.

Its rank is not full. Its pseudo inverse brings the $C(A_3)$ to the row space

$$A_3^+ = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}^+ = \frac{v_1 u_1^T}{\sigma_1} = \frac{1}{10}\begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix}$$

$\boxed{7.1}$

2. We usually think that the identity matrix $I$ is as simple as possible. But why is $I$ completely incompressible ?
Draw a rank-5 flag with a cross.

Ans: $\sigma_i = 1$ for all $i$

$\Longrightarrow$ we can't leave out any of the terms $u_i \cdot v_i^T$ without making an error of size 1.

3- These flags have rank 2. Write A & B in any way as $u_1 v_1^T + u_2 v_2^T$.

$$A_{sweden} = A_{Finland} = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 2 & 2 & 2 & 2 \\ 1 & 2 & 1 & 1 \end{bmatrix}$$

$$B_{Benin} = \begin{bmatrix} 1 & 2 & 2 \\ 1 & 3 & 3 \end{bmatrix}$$

Ans:

$$BB^T = \begin{bmatrix} 1 & 2 & 2 \\ 1 & 3 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 9 & 13 \\ 13 & 19 \end{bmatrix}$$

$$\sigma_1^2 = 14 + \sqrt{194}, \quad \sigma_2^2 = \frac{2}{14 + \sqrt{194}}.$$

$$\sigma_2 \approx \frac{1}{\sqrt{14}}. \implies B \text{ is compressible.}$$

The good row $v_1$ and column $u_1$ are eigenvectors of $B^T B$ and $BB^T$.

$\boxed{7.2}$

(2/3) Find $A^TA$ and $V$ & $\Sigma$ and $u_i = \frac{Av_i}{\sigma_i}$ and the full SVD,

$$A = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} = U \Sigma V^T$$

Show that $AA^T$ is diagonal. Its eigenvectors $u_1, u_2$ are ________. Its eigenvalues are $\sigma_1^2, \sigma_2^2$ are ________. The rows of 'A' are orthogonal but they are not ________.
So the columns of 'A' are not orthogonal.

Ans. $A^TA = \begin{bmatrix} 2 & -1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$

$AA^T = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 2 & -1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix}$

$\sigma_1^2 = 8, \ \sigma_2^2 = 2$

$A^TA$ with eigenvectors in $V = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$

$Av = \sigma_i u_i$

$u_1 = \frac{Av_1}{\sigma_1} = \frac{1}{2\sqrt{2}}\begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix}\frac{1}{\sqrt{2}} = \frac{1}{4}\begin{bmatrix} 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$u_2 = \frac{Av_2}{\sigma_2} = \frac{1}{\sqrt{2}}\begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \end{bmatrix}\frac{1}{\sqrt{2}} = \frac{1}{2}\begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$$A = U\Sigma V^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \frac{1}{\sqrt{2}}$$

$AA^T = \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix}$ is a diagonal matrix. So its eigenvectors $(1,0)$ and $(0,1)$ go in $U = I$.

The rows of $A$ are orthogonal, but not orthonormal.

$\therefore$ $A^T A$ is not diagonal and $V$ is not $I$.

7. If $(A^T A)v = \sigma^2 v$, multiply by $A$. Move the parentheses to get $(AA^T)Av = \sigma^2 (Av)$.

*→ If $v$ is an eigenvector of $A^T A$, then ______ is an eigenvector of $AA^T$.

Ans: $(A^T A)v = \sigma^2 v = \lambda v$

$$(AA^T)Av = A(A^TA)v = A\lambda v = \lambda Av$$

$\Longrightarrow$ $Av$ is an eigenvector of $AA^T$ with the same eigenvalue $\lambda$.

10. (a) Why is the trace of $A^TA$ equals to the sum of all $a_{ij}^2$ ?

(b) For every rank-1 matrix, why is $\sigma_1^2 =$ sum of all $a_{ij}^2$ ?

Ans: (a) Every diagonal entry of $A^TA$ is the sum of $a_{ij}^2$ down one column.

So, the trace is the sum down all columns.

$\longrightarrow$ $tr(A^TA) =$ sum of all $a_{ij}^2$

11. SVD of Fibonacci matrix, $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$

Ans: $A^T A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = AA^T$

$\sigma_1^2 = \frac{3+\sqrt{5}}{2}$ & $\sigma_2^2 = \frac{3-\sqrt{5}}{2}$

$\sigma_1 = \frac{\sqrt{5}+1}{2}$ , $\sigma_2 = \frac{\sqrt{5}-1}{2}$

11. SVD of Fibonacci matrix, $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$

Ans: $A^T A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = A A^T$

$\sigma_1^2 = \dfrac{3+\sqrt{5}}{2}$ & $\sigma_2^2 = \dfrac{3-\sqrt{5}}{2}$

$\sigma_1 = \dfrac{\sqrt{5}+1}{2}$, $\sigma_2 = \dfrac{\sqrt{5}-1}{2}$

15. Construct the matrix with rank-1 that has $Av = 12u$ for $v = \frac{1}{2}(1,1,1,1)$ and $u = \frac{1}{3}(2,2,1)$. Its only singular value is $\sigma_1 =$ ______

Ans: $Av = \sigma_1 u = 12u$

$A = uw^T$ for some $w$, & $u \in C(A)$

and $w \in C(A^T)$

$w = kv$.

$\Longrightarrow \underline{\underline{A = 12uv^T}}$

16. Suppose 'A' has orthogonal columns $w_1, \ldots, w_n$ of lengths $\sigma_1, \ldots, \sigma_n$. What are $U, \Sigma, V$ in the SVD.?

Ans: $A^T A =$ diagonal matrix with diagonal entries $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$.

$V = I$

$Av_i = \sigma_i u_i$

$A = X \Sigma I = (A\Sigma^{-1}) \Sigma I$

17. Suppose $A$ is $2\times 2$ symmetric matrix with unit eigenvectors $u_1$ and $u_2$. If its eigenvalues are $\lambda_1 = 3$ and $\lambda_2 = -2$, what are the matrices $U, \Sigma, V^T$ in its SVD?

Ans: $A^T = A$.

$$Au_1 = 3u_2$$
$$Au_2 = -2u_2$$

$$A^TA = AA^T = A^2$$

$$\sigma_i^2 = \lambda_i^2$$

$$\sigma_1 = |\lambda_1| = 3 \quad , \quad \sigma_2 = |\lambda_2| = |-2| = 2$$

$$\boxed{\Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}}$$

$\cancel{A^2u_1 = 9u_1}$ & $\cancel{A^2u_2 = 4u_2}$

$$AA^Tu_i = A^2u_i = \sigma_i^2 u_i$$

$$A^TAv_i = A^2v_i = \sigma_i^2 v_i$$

$u_1 = v_1$ & $u_2 = -v_2$ (due to ~~the sign~~ $\lambda_2 = -2$)

18. If $A = QR$, with an orthogonal matrix $Q$, the SVD of $A$ is almost the same as the SVD of $R$. Which of the 3 matrices $U, \Sigma, V$ is changed because of $Q$?

Ans: Let, $R = U\Sigma V^T$

$$A = QR = Q(U\Sigma V^T) = (QU)\Sigma V^T$$

$$(QU)(QU)^T = QUU^TQ^T = QQ^T = I,$$

$QU$ is ~~also~~ orthogonal.

SVD of $A$ is $(QU)\Sigma V^T$

19. Suppose $A$ is invertible (with $\sigma_1 > \sigma_2 > 0$). Change '$A$' by as small a matrix as possible to produce a singular matrix $A_0$.

From $A = \begin{bmatrix} u_1 & v_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & \\ & \sigma_2 \end{bmatrix} \begin{bmatrix} v_1 & v_2 \end{bmatrix}^T$

find the nearest $A_0$.

Ans: The smallest change in 'A' is to set its smallest singular value $\sigma_2$ to zero.

Q4. Find the max. of $\dfrac{x^T S x}{x^T x} = \dfrac{3x_1^2 + 2x_1x_2 + 3x_2^2}{x_1^2 + x_2^2}$
(a)

ILA (a) What matrix is S.

Ans: $\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 3x_1^2 + 2x_1x_2 + 3x_2^2$

$\lambda_1 = 4, \; \lambda_2 = 2$

max. of $\dfrac{x^T S x}{x^T x} = 4.$

(b) Find the max of $\dfrac{(x_1 + 4x_2)^2}{x_1^2 + x_2^2}$.
For what matrix A is this $\dfrac{\|Ax\|^2}{\|x^2\|}$ ?

Ans: $A_{1\times 2} = \begin{bmatrix} 1 & 4 \end{bmatrix}$

max. is $\sigma_1^2(A) = 17$ since
$AA^T = \begin{bmatrix} 1 & 4 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix} = [17]$.

25. What are the min. values of the ratios $\frac{x^T S x}{x^T x}$ and $\frac{\|Ax\|^2}{\|x\|^2}$ ? We should take $x$ to be which eigenvectors of $S$ ? Should $x$ always be an eigenvector of $A$ ?

Ans:

The min. value of $\frac{x^T S x}{x^T x}$ is the smallest eigenvalue of $S$. The eigenvector is the minimizing $x$.

---

ILA ①

$$AB = A\left[\vec{b_1}\ \vec{b_2} \cdots \vec{b_n}\right] = \left[A\vec{b_1}\ \ A\vec{b_2} \cdots A\vec{b_n}\right]$$

$$A\vec{v} = v_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + \cdots + v_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

$\Rightarrow A\vec{v}$ is a linear combination of columns of $A$.

$\Rightarrow$ Every column of $AB$ is a linear combination of columns of $A$.

$\Rightarrow$ $\underline{\underline{C(AB) \subset C(A)}}$

27. all matrices $A$ with a given column space in $\mathbb{R}^m$ and a given row space in $\mathbb{R}^n$. Suppose $c_1, \ldots, c_r$ and $b_1, \ldots, b_r$ are bases for those 2 spaces. Make them columns of $C$ & $B$. The goal is to show that $A$ has this form:
'$A$' $= CMB^T$ for an $r \times r$ invertible matrix $M$.

Ans: $$A = U \Sigma V^T$$

columns of $U$ are a basis for $C(A)$.
& so are the columns of $C$. $\to$ $\underline{\underline{C(U) = C(C)}}$
$\Longrightarrow$ $U = CF$ for some $\underline{\text{invertible}}$ $F$

Similarly,
the columns of $V$ are a basis for the row space of '$A$' & so are the columns of $B$, $\longrightarrow$ $V = BG$ for some invertible $r \times r$ matrix $G$.

then,

$$A = U \Sigma V^T = (CF) \Sigma (BG)^T = C \left( F \Sigma G^T \right) B^T = CMB^T$$

and $M = F \Sigma G^T$ is $r \times r$ & invertible

$\boxed{7.4}$

1. $A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$, rank-1 matrix

$AA^T u_i = \sigma_i^2 u_i$ &

Ans: $A^T A v_i = \sigma_i^2 v_i$

$$AA^T = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}\begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix} = \begin{bmatrix} 5 & 15 \\ 15 & 45 \end{bmatrix} = 5\begin{bmatrix} 1 & 3 \\ 3 & 9 \end{bmatrix}$$

$$(1-\lambda)(9-\lambda) - 9 = 0 = \lambda^2 - 10\lambda = 0 \Rightarrow \lambda(\lambda - 10) = 0.$$

$\lambda_2 = 0, \; \lambda_1 = 50$

$v_2 = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 \\ -1 \end{bmatrix}, \; v_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} = U\Sigma V^T = \frac{1}{\sqrt{10}}\begin{bmatrix} 1 & -3 \\ +3 & 1 \end{bmatrix}\begin{bmatrix} 5\sqrt{2} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & +2 \\ -2 & 1 \end{bmatrix}\frac{1}{\sqrt{5}}$$

$v_1$ for row space, $v_2$ for null space

$u_1$ for column space, $u_2$ for $N(A^T)$

In this case,

all matrices with those 4 subspaces are multiples $cA$; since the subspaces are just lines.

$$A = U\Sigma V^T = (UV^T)(V\Sigma V^T) = QS$$

$$Q = UV^T = \frac{1}{\sqrt{10}}\begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}\frac{1}{\sqrt{5}}$$

$$= \frac{1}{\sqrt{50}}\begin{bmatrix} -5 & 5 \\ -5 & 5 \end{bmatrix} = \frac{1}{5\sqrt{2}}\begin{bmatrix} -5 & 5 \\ -5 & -5 \end{bmatrix}$$

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

$$= \frac{1}{\sqrt{50}}\begin{bmatrix} 7 & -1 \\ 1 & 7 \end{bmatrix}$$

$$S = V\Sigma V^T = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 5\sqrt{2} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}\frac{1}{\sqrt{5}}$$

$$= \frac{1}{5}\begin{bmatrix} 5\sqrt{2} & 10\sqrt{2} \\ 10\sqrt{2} & 20\sqrt{2} \end{bmatrix} = \frac{1}{5\sqrt{2}}\begin{bmatrix} 10 & 20 \\ 20 & 40 \end{bmatrix}$$

$$= \frac{1}{\sqrt{50}}\begin{bmatrix} 10 & 20 \\ 20 & 40 \end{bmatrix}$$

$$A^+ = V\Sigma^+U^T$$

$$= \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} \frac{1}{5\sqrt{2}} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 3 \\ -3 & 1 \end{bmatrix}\frac{1}{\sqrt{10}}$$

$$= \frac{1}{\sqrt{50}}\begin{bmatrix} 1/5\sqrt{2} & 3/5\sqrt{2} \\ \sqrt{2}/5 & 3\sqrt{2}/5 \end{bmatrix} = \frac{1}{50}\begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$$

$$P_{C(A)} = AA^+ = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \cdot \frac{1}{50}\begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$$

$$= \frac{1}{50}\begin{bmatrix} 5 & 15 \\ 15 & 45 \end{bmatrix} = \begin{bmatrix} 0.1 & 0.3 \\ 0.3 & 0.9 \end{bmatrix}$$

$\frac{1}{\sqrt{5}}$

$$P_{C(A^T)} = A^+A = \frac{1}{50}\begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$$

$$= \frac{1}{50}\begin{bmatrix} 10 & 20 \\ 20 & 40 \end{bmatrix} = \begin{bmatrix} 0.2 & 0.4 \\ 0.4 & 0.8 \end{bmatrix}$$

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

$A$ is invertible $\Longleftarrow |A| = 6 \neq 0$.

Ans:

$$A = \begin{bmatrix} 3 & 3 \\ -1 & 1 \end{bmatrix} = U\Sigma V^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix}\begin{bmatrix} 1 & +1 \\ -1 & 1 \end{bmatrix}\frac{1}{\sqrt{2}}$$

$$A = U\Sigma V^T = (U\Sigma U^T)(UV^T) = KQ$$

where,

$$K = U\Sigma U^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix}$$

$$Q = UV^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & +1 \\ -1 & 1 \end{bmatrix}\frac{1}{\sqrt{2}} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

$$A^+ = V\Sigma^+ U^T = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} \frac{1}{3\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$= \frac{1}{\sqrt{2}}\begin{bmatrix} \frac{1}{3\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{3\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1/6 & -1/2 \\ 1/6 & 1/2 \end{bmatrix} = A^{-1}$$

10. $A = \begin{bmatrix} 3 & 4 & 0 \end{bmatrix}_{1\times 3}$

Ans: $AA^T = \begin{bmatrix} 3 & 4 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 25 \end{bmatrix}$

$\sigma_{x_i}^2 = 25, 0, 0.$

$$A^T A = \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix} \begin{bmatrix} 3 & 4 & 0 \end{bmatrix} = \begin{bmatrix} 9 & 12 & 0 \\ 12 & 16 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$V_1 = \frac{1}{5}\begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix}, \quad V_2 = \frac{1}{5}\begin{bmatrix} -4 \\ 3 \\ 0 \end{bmatrix}, \quad V_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 \end{bmatrix}_{1\times 1} \begin{bmatrix} 5 & 0 & 0 \end{bmatrix}_{1\times 3} \begin{bmatrix} 3 & 4 & 0 \\ -4 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}_{3\times 3} \frac{1}{5} = U \Sigma V^T$$

$$\cancel{A^+ = V \Sigma^+ U^T = \frac{1}{5}\begin{bmatrix} 3 & -4 & 0 \\ 4 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} 1/5 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix} =}$$

$$A^+ = V\Sigma^+U^T = \frac{1}{5}\underbrace{\begin{bmatrix} 3 & -4 & 0 \\ 4 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}}_{3\times 3}\underbrace{\begin{bmatrix} 1/5 \\ 0 \\ 0 \end{bmatrix}}_{3\times 1}\begin{bmatrix} 1 \end{bmatrix}$$

$$A^+ = \frac{1}{25}\begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix}$$

$$A^+A = \frac{1}{25}\begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix}\begin{bmatrix} 3 & 4 & 0 \end{bmatrix} = \frac{1}{25}\begin{bmatrix} 9 & 12 & 0 \\ 12 & 16 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0.36 & 0.48 & 0 \\ 0.48 & 0.64 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$AA^+ = \begin{bmatrix} 3 & 4 & 0 \end{bmatrix}\begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix}\frac{1}{25} = \frac{1}{25}\begin{bmatrix} 25 \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix}$$

12. What is the only $2\times 3$ matrix that has no pivots and no singular values ?
What is $\Sigma$ for that matrix ?

$A^+$ is the zero matrix, but what is its shape ?

Ans

Ans: The zero matrix has no pivots or singular values.

$\Sigma$ = $2\times3$ zero matrix

Pseudo inverse is $3\times2$.

13. If $\det(A)=0$, why is $\det(A^+)=0$? If $A$ has rank $r$, why does $A^+$ have rank $r$?

Ans: $A = U\Sigma V^T$

$U$ & $V$ are orthonormal matrices.

$|U| \neq 0, |V| \neq 0$

$\therefore |A| = 0 \Longrightarrow |\Sigma| = 0$

$\Longrightarrow \text{rank}(A) < n$

$$\boxed{C(A^+) = C(A^T) \ \& \ N(A^+) = N(A^T)}$$

$\Longrightarrow \text{rank}(A^T) = \text{rank}(A) < n$

$\therefore$ ~~$|A^+| \neq$~~ " $\det(A^+) = 0$

14. The matrix '$A$' transforms the circle of unit vectors $\|x\|=1$ into an ellipse of vectors $y = Ax$. The reason is that $x = A^{-1}y$ and the vectors with $\|A^{-1}y\|=1$ do lie on an ellipse:

$$\|A^{-1}y\|^2 = 1 \implies (A^{-1}y)^T(A^{-1}y) = 1$$

$$y^T(A^{-1})^T A^{-1} y = 1$$

$$y^T (AA^T)^{-1} y = 1 \Rightarrow y^T S y = 1$$

$(A^TA)^{-1}$ is symmetric +ve definite.

---

For $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \implies AA^T = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$

$$(AA^T)^{-1} = \frac{1}{9}\begin{bmatrix} 5 & -4 \\ -4 & 5 \end{bmatrix}$$

$\therefore$ the ellipse $\|A^{-1}y\|^2=1$ of outputs $y = Ax$ has equation $5y_1^2 - 8y_1y_2 + 5y_2^2 = 9$.

14\. The matrix 'A' transforms the circle of unit vectors $\|x\|=1$ into an ellipse of vectors $y=Ax$. The reason is that $x=A^{-1}y$ and the vectors with $\|A^{-1}y\|=1$ do lie on an ellipse:

$$\|A^{-1}y\|^2 = 1 \Longrightarrow (A^{-1}y)^T(A^{-1}y) = 1$$

$$y^T(A^{-1})^T A^{-1} y = 1$$

$$y^T(AA^T)^{-1} y = 1 \Rightarrow y^T S y = 1$$

$(A^TA)^{-1}$ is symmetric +ve definite.

For $A=\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix} \Longrightarrow AA^T = \begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix} = \begin{bmatrix}5 & 4\\ 4 & 5\end{bmatrix}$

$$(AA^T)^{-1} = \frac{1}{9}\begin{bmatrix}5 & -4\\ -4 & 5\end{bmatrix}$$

$\therefore$ the ellipse $\|A^{-1}y\|^2=1$ of outputs $y=Ax$ has equation $5y_1^2 - 8y_1y_2 + 5y_2^2 = 9$.

15. All matrices have rank-1. The vector $b$ is $(b_1, b_2)$:

$$A = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix},\ AA^T = \begin{bmatrix} 8 & 4 \\ 4 & 2 \end{bmatrix},\ A^TA = \begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix},\ A^+ = \begin{bmatrix} 0.2 & 0.1 \\ 0.2 & 0.1 \end{bmatrix}$$

(a) The eq$^n$. $A^TA\hat{x} = A^Tb$ has many solutions because $A^TA$ is ________________

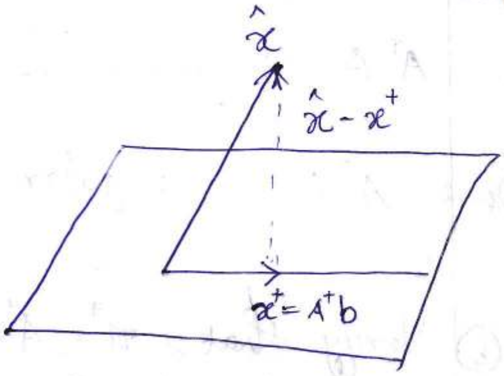
Ans: $A^TA$ is singular

(b) Verify that $x^+ = A^+b = (0.2b_1 + 0.1b_2,\ 0.2b_1 + 0.1b_2)$ solves $A^TAx^+ = A^Tb$.

Ans:

16. The vector $x^+ = A^+ b$ is the shortest possible solution to $A^T A \hat{x} = A^T b$.

Reason :




The difference $\hat{x} - x^+$ is in the null space of $A^T A$.

$$\hat{x} - x^+ \in N(A^T A)$$

$$\hat{x} - x^+ \in N(A)$$

$$\hat{x} - x^+ \perp x^+$$

$$\Longrightarrow \boxed{\|\hat{x}\|^2 = \|x^+\|^2 + \|\hat{x} - x^+\|^2}$$

17. Every 'b' in $\mathbb{R}^m$ is $p+e$. This is the column space part + left nullspace part.

Every 'x' in $\mathbb{R}^n$ is $x^+ + x_n$. This is the row space part + nullspace part. Then,

$AA^+p =$ ____________

$AA^+e =$ ____________

$A^+Ax^+ =$ ____________

$A^+Ax_n =$ ____________

Ans: $AA^+p = P_{C(A)}p = p$

$AA^+e = 0$

$A^+Ax^+ = P_{C(A^T)}x^+ = x^+$

$A^+Ax_n = 0.$

21. From $A$ & $A^+$ show that $A^+A$ is correct and $(A^+A)^2 = A^+A$ = projection.

Ans: $A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$

$A^+ = \sum_{i=1}^{r} \frac{v_i u_i^T}{\sigma_i}$

$$A^+A = \left(\sum_{i=1}^{r} \frac{v_i u_i^T}{\sigma_i}\right)\left(\sum_{i=1}^{r} \sigma_i u_i v_i^T\right)$$

$$= \sum_{i=1}^{r} v_i u_i^T u_i v_i^T$$ since $u_i^T u_j = 0$ when $i \neq j$

$u_i^T u_i = 1$ for $i = j$

$$\boxed{A^+A = \sum_{i=1}^{r} v_i v_i^T}$$

$$\boxed{AA^+ = \sum_{i=1}^{r} u_i u_i^T}$$

22. Each pair of singular vectors $v$ and $u$ has $Av = \sigma u$ and $A^T u = \sigma v$. Show that the double vector $\begin{bmatrix} v \\ u \end{bmatrix}$ is an eigenvector of the symmetric block matrix $M = \begin{bmatrix} 0 & A^T \\ A & 0 \end{bmatrix}$.

The SVD of 'A' is equivalent to the diagonalization of that symmetric matrix M.

Ans: $M \begin{bmatrix} v \\ u \end{bmatrix} = \begin{bmatrix} 0 & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ u \end{bmatrix} = \begin{bmatrix} A^T u \\ Av \end{bmatrix} = \begin{bmatrix} \sigma v \\ \sigma u \end{bmatrix} = \sigma \begin{bmatrix} v \\ u \end{bmatrix}$

$\Longrightarrow \begin{bmatrix} v \\ u \end{bmatrix}$ is an eigenvector.

& singular values of 'A' are eigenvalues of this block matrix.