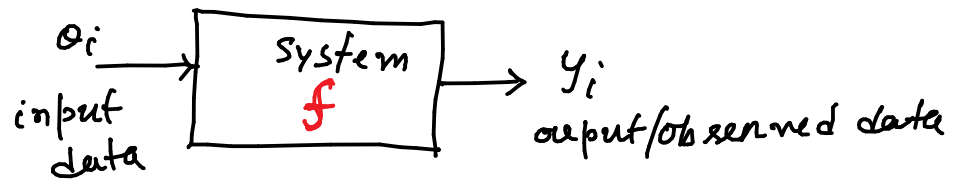


NL/07/11/25



From an experiment, we have collected following data set:

x_i	y_i
x_1	y_1
x_2	y_2
x_3	y_3

input data $\downarrow$ $\uparrow$ output data on observed data

x_i	y_i
0	1
1	2
3	-1
4	1

Let us try to learn the model

Determine f

which is compatible with the collected data set

We need to find f s.t. we have

$$\text{output data } y_i = f(x_i)$$

$$\text{or at least } y_i \approx f(x_i)$$

← may not be possible!

- We construct (learn) a model f from the data!

- Denote the output of model $f(x_i)$

as $\hat{y}_i = f(x_i)$ which may not equal to the actual observation data y_i .

- For linear model:

Let $\hat{y}_i = f(\theta_i) = a_1 \theta_i + a_0$

Collected data

θ_i	y_i
1	-1
2	1
0	2
-1	3

$$\begin{aligned}
 -1 &= a_1 \cdot 1 + a_0 \\
 1 &= a_1 \cdot 2 + a_0 \\
 2 &= a_1 \cdot 0 + a_0 \\
 3 &= a_1 \cdot (-1) + a_0
 \end{aligned}$$

$\uparrow$ $\uparrow$
 y_i θ_i

- The actual output y_i may not equal to predicted data $\hat{y}_i$ obtained from

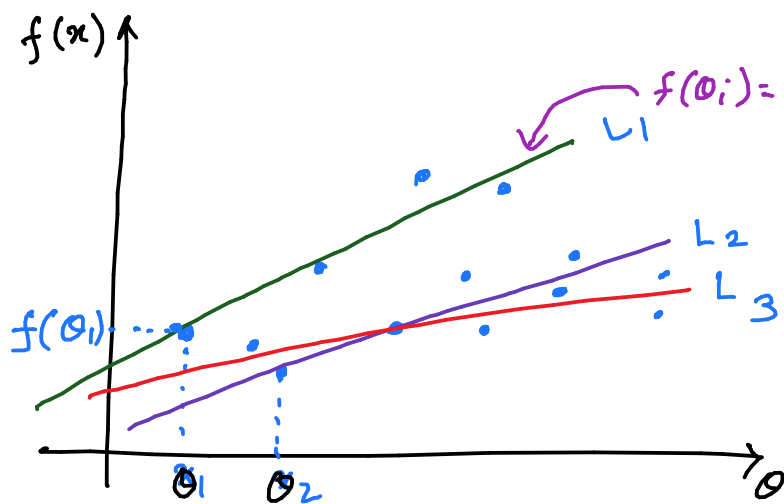
$$\hat{y}_i = a_1 \theta_i + a_0$$

↓ can be represented as

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 0 & 1 \\ -1 & 1 \end{bmatrix}}_A \underbrace{\begin{bmatrix} a_1 \\ a_0 \end{bmatrix}}_x = \underbrace{\begin{bmatrix} -1 \\ 1 \\ 2 \\ 3 \end{bmatrix}}_b$$

$$Ax = b$$

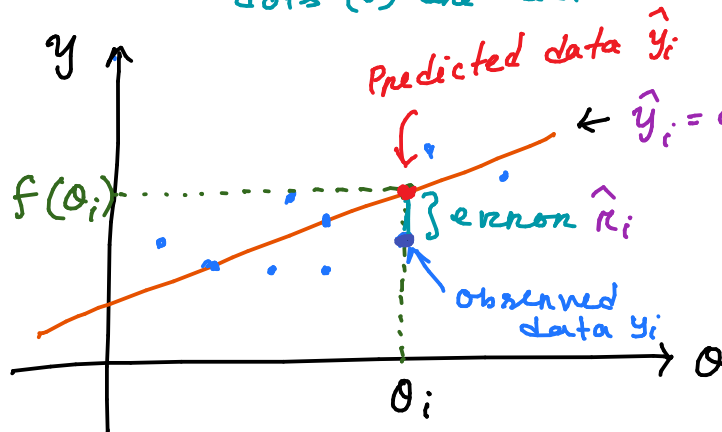
Find x to get the model.



(points on the line are predicted data & other dots (•) are actual output data)

corresponding to the data collected we have $L_1, L_2, L_3 \dots$

which line is correct?



Predicted data $\hat{y}_i$

$\hat{y}_i = a_1 \theta_i + a_0$ • what even

line we choose, some

points may lie on the line &

some may not.

Error between observed data (y_i)

& predicted data $f(\theta_i) = a_1 \theta_i + a_0$

Residue

$$r_i = y_i - f(\theta_i) = (y_i - \hat{y}_i)$$

• We consider the line which is

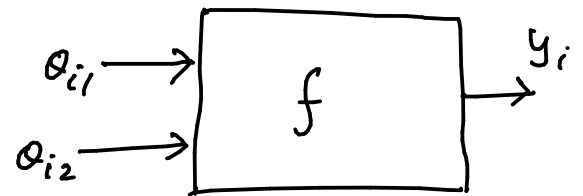
obtained by $\rightarrow \min \left(|r_1|^2 + |r_2|^2 + \dots + |r_n|^2 \right)$

The obtained line would be the

"best fit line" with the

The problem is called "Least Square problem". data.

θ_{i1}	θ_{i2}	y_i
1	0	2
3	1	4
-1	0	0
-2	1	-1



Model

$$f(\theta_{i1}, \theta_{i2}) = a_1 \theta_{i1} + a_2 \theta_{i2} + a_0 \quad \leftarrow \text{outputs will be } \hat{y}_i$$

$$2 = a_1 \cdot 1 + a_2 \cdot 0 + a_0$$

$$4 = a_1 \cdot 3 + a_2 \cdot 1 + a_0$$

$$0 = a_1 \cdot (-1) + a_2 \cdot 0 + a_0$$

$$-1 = a_1 \cdot (-2) + a_2 \cdot 1 + a_0$$

y_i $\uparrow$

$\downarrow$

$$\begin{bmatrix} 1 & 0 & 1 \\ 3 & 1 & 1 \\ -1 & 0 & 1 \\ -2 & 1 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_0 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 0 \\ -1 \end{bmatrix}$$

$\downarrow$
observed data

Since we always have large no. of experimental data $\downarrow$

A is a tall matrix

$Ax = b \rightarrow$
Find Θ to get the model

$$Ax = b$$

Residue $r = b - Ax \rightarrow \|r\|_2^2 = \|b - Ax\|_2^2$

Objective is to

$$\text{minimize } \|r\|_2^2$$

$$\|r\|_2^2 = |r_1|^2 + |r_2|^2 + \dots + |r_n|^2$$

$\downarrow$
to obtain the model i.e. Θ

- One can solve for x directly for $Ax = b$ if the observed data (b) belongs to the column space of matrix A .

- Since A is a tall matrix, direct solving of $Ax = b$ for x may not always be possible!



Go with solving the optimization problem to find x .



Unconstrained
convex optimization
problem

$$\left\{ \begin{array}{l} \min \| \hat{r} \|_2^2 \leftarrow \text{cost fun}^n \\ = \min \| b - Ax \|_2^2 \\ \text{for all } x \in \mathbb{R}^n \end{array} \right.$$



Coefficients of
the model

Necessary & Sufficient condition for such optimization problem:



$$\boxed{\nabla f(\bar{x}) = 0}$$

← gradient must be zero at optimal value $\bar{x}$.

$f(x)$	$\nabla f(x)$
Ax	A^T
$x^T A$	A
$x^T x$	$2x$
$x^T A x$	$Ax + A^T x \checkmark$
$b^T A x$	$A^T b$

Modify the cost funⁿ $\rightarrow \frac{1}{2} \|\hat{r}\|_2^2$

$$\begin{aligned}
 \frac{1}{2} \|\hat{r}\|_2^2 &= \frac{1}{2} [(Ax - b)^T (Ax - b)] \\
 &= \frac{1}{2} [(x^T A^T - b^T) (Ax - b)] \\
 &= \frac{1}{2} [x^T A^T A x - \underbrace{x^T A^T b}_{\text{scalars}} - \underbrace{b^T A x}_{\text{scalars}} + b^T b] \\
 &= \frac{1}{2} [x^T A^T A x - 2b^T A x + b^T b]
 \end{aligned}$$

$$\begin{aligned}
 \nabla f(x) &= \frac{1}{2} [2A^T A x - 2A^T b] \\
 &= A^T A x - A^T b
 \end{aligned}$$

At Optimal value $\bar{x} \rightarrow \nabla f(\bar{x}) = 0$

$$\nabla f(\bar{x}) = 0$$

$$A^T A \bar{x} - A^T b = 0$$

$$\Rightarrow A^T A \bar{x} = A^T b$$

$$\Rightarrow \bar{x} = (A^T A)^{-1} A^T b$$

A must be full column rank.

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

$$\bar{x} = A^+ b \quad \text{Optimal solution}$$

$$\bar{x} = \begin{bmatrix} a_{n-1} \\ a_{n-2} \\ \vdots \\ a_1 \\ a_0 \end{bmatrix}$$

model coefficients

NL/11/11/25



Let f be in the form $a_1 x_i + a_0$

↓

Find best fit line $f(x) = a_1 x_i + a_0$
with the input/output data

"From the data we are learning the function f ".

Linear Regression Problem
or
Least Square Problem

To identify the best fit line
we define residue

$$r_i = y_i - f(x_i)$$

Combining all data
 $Ax = b$

↑
observed data

↑
Predicted data

Find x s.t. $(|r_1|^2 + |r_2|^2 + \dots + |r_n|^2)$
is minimum

For a given set of data, one can write it in

$$Ax = b$$

$$\min_x \|r\|_2^2 = \min_x \|b - Ax\|_2^2$$

↓

For given $Ax = b$

$R(A)$:
Image space
of A

solⁿ exists
(if $b \in R(A)$)

solⁿ may not exist
(sim A is tall matrix)

↓
form for x

↓
obtain Least Square solⁿ

||

$$\min \|b - Ax\|_2^2$$

↓
Least Square solⁿ

↓

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

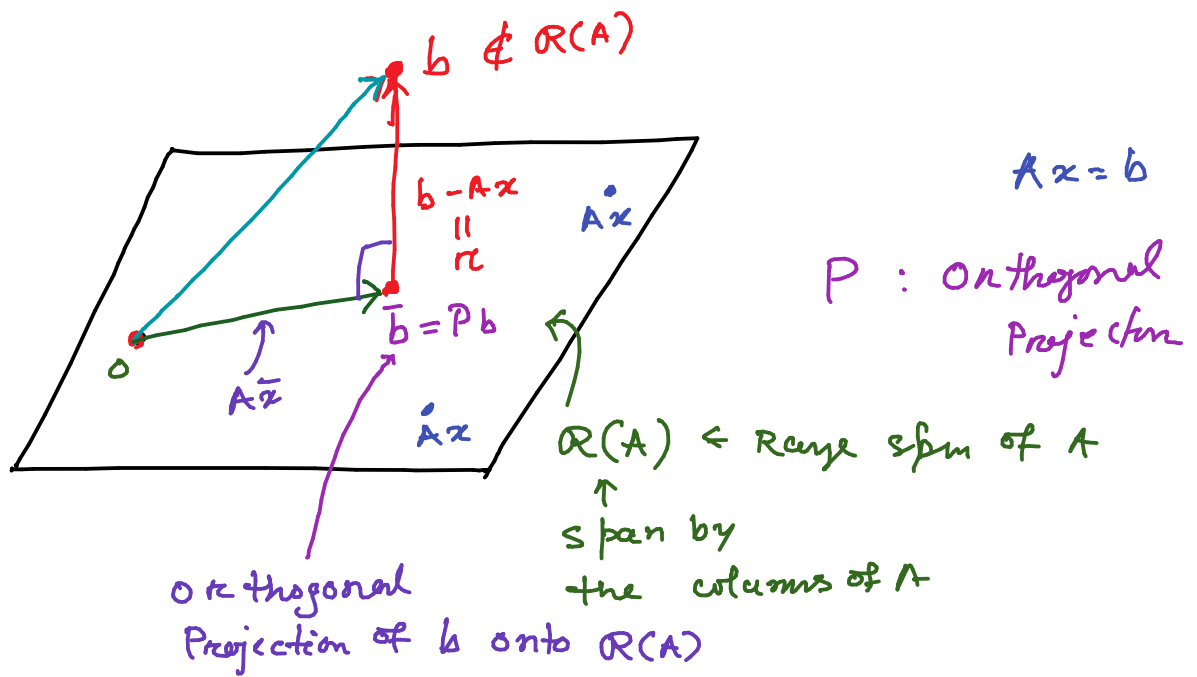
$$\bar{x} = A^+ b$$

optimal
value

A is full column
rank
 $\Rightarrow A^T A$ is invertible

A^+ Pseudo-inverse of A

$$\downarrow \\ (A^T A)^{-1} A^T$$



We need the minimum r

$$\min \|r\|_2^2 = \min_x \|b - Ax\|$$

$$= \|b - \bar{b}\| \quad \left(\begin{array}{l} \text{Distance bet}^{\circ} \\ \text{a subspace \& a point} \end{array} \right)$$

$$\text{where } \bar{b} = \underset{\uparrow}{P} b$$

Orthogonal Projection that projects b onto $R(A)$.

- For a full col. rank matrix the orthogonal

Projection that projects b onto

$$R(A) \rightarrow P = AA^+$$

$$= A \left[(A^T A)^{-1} A^T \right]$$

Now recall the optimal value .

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

$$A\bar{x} = \underbrace{A(A^T A)^{-1} A^T}_{= P} b$$

$$\Rightarrow A\bar{x} = Pb = \bar{b}$$

$$A\bar{x} \in \mathcal{R}(A)$$

- Hence, the optimal value of least square problem can be obtained from the observed data vector 'b' by projecting it Orthogonally onto the image space of A i.e. $\mathcal{R}(A)$



Find Orthogonal Projection P that projects b onto $\mathcal{R}(A)$

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

↓ compute

$$\bar{b} = Pb = A(A^T A)^{-1} A^T b$$

↓ solve for x

$$Ax = \bar{b} \quad \leftarrow \text{Solvable since}$$

Solution will be $\bar{x} = A^+ b$ $\bar{b} \in \mathcal{R}(A)$.

Verify

$$A\bar{x} = AA^+ b = Pb = \bar{b}$$

→ Minimum Norm Solution
of a System of Linear Equations.

Optimizant

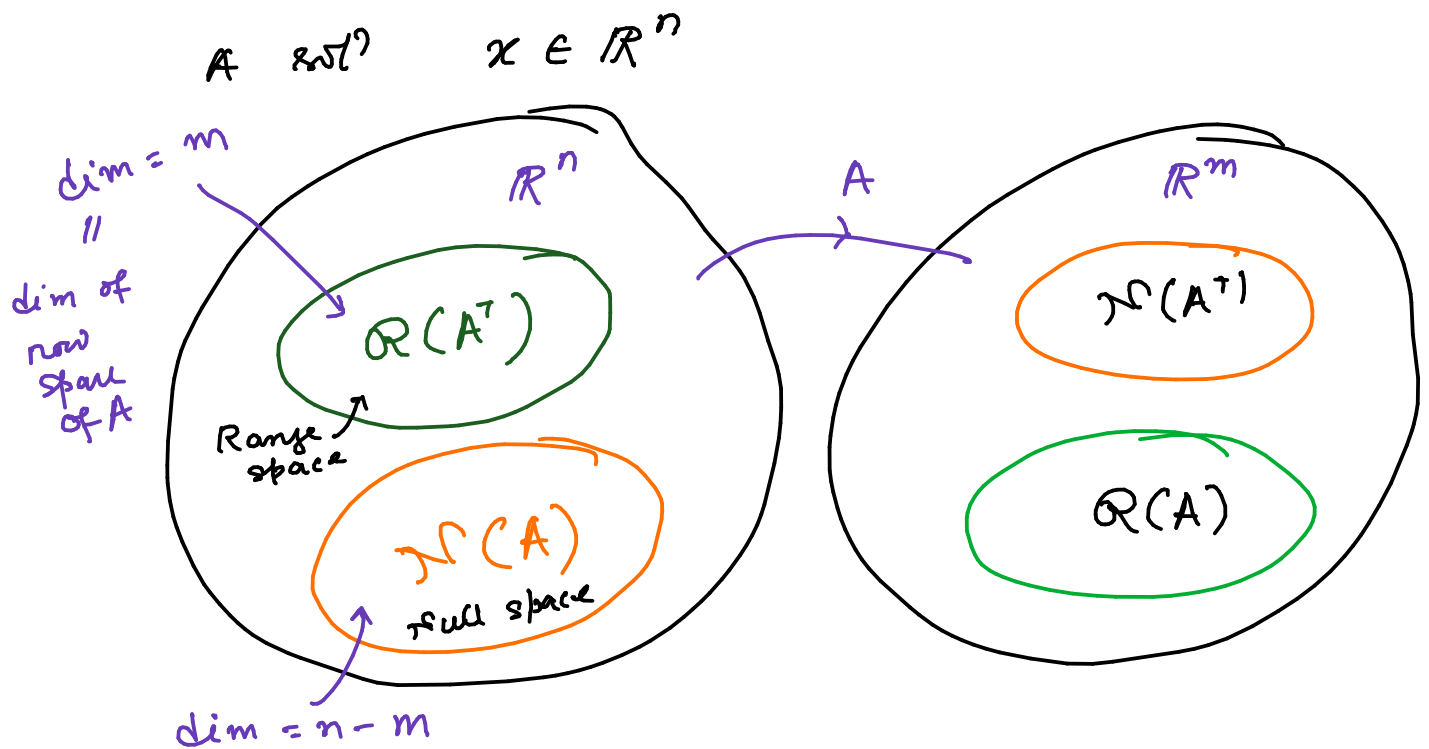
$$\left\{ \begin{array}{l} \min \|x\| \\ \text{s.t. } Ax = b \end{array} \right.$$

when $A \in \mathbb{R}^{m \times n}$ is full
row rank
matrix.

- $[A] \leftarrow$ square (Unique solⁿ if A is non-singular)
- $[A] \leftarrow$ Tall matrix (Solⁿ may not exist)
- $[A] \leftarrow$ Fat matrix (Possibility of infinitely many solⁿ)

$$Ax = b \quad \text{with} \quad A \text{ full row rank}$$

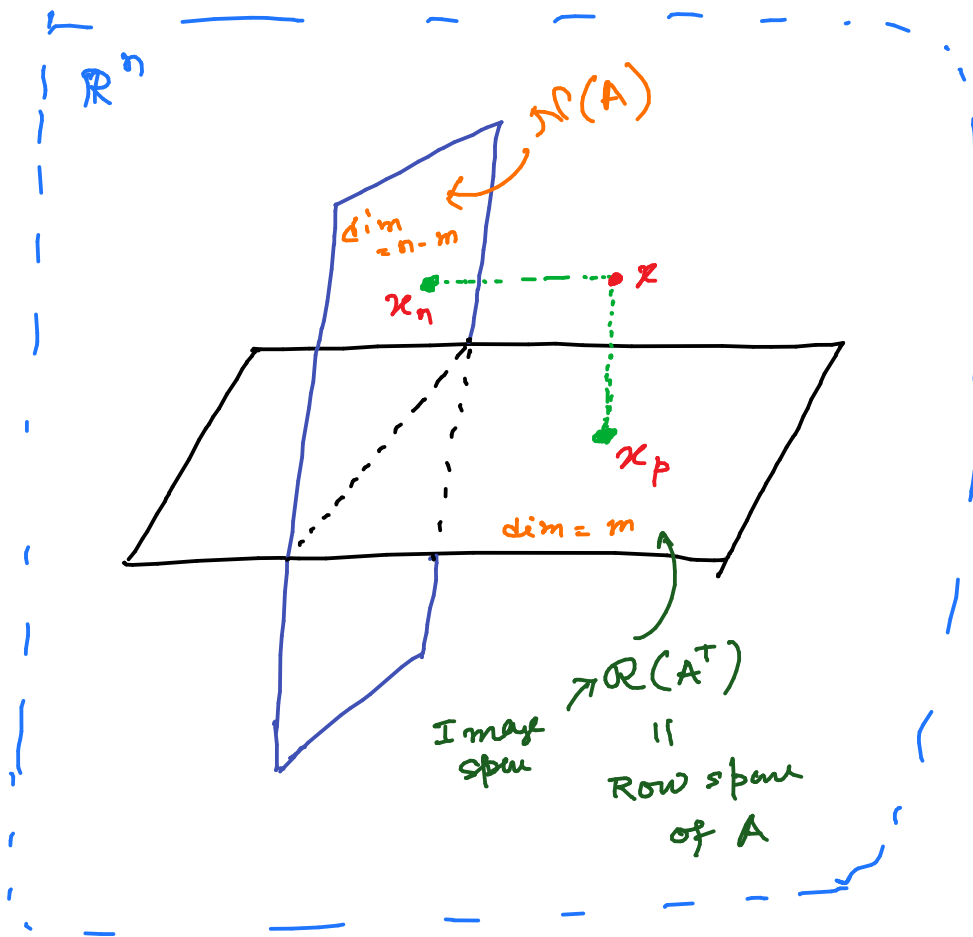
$$A \in \mathbb{R}^{m \times n}$$



$$\text{Rel}^n \text{ bet } R(A^T) \text{ \& } N(A)$$

$$N(A) \oplus R(A^T)$$

$$[R(A^T)]^\perp = N(A)$$



For $Ax = b$
 $x \in \mathbb{R}^n$

$$Px = x_p$$

P is an orthogonal projection.

$x \in \mathbb{R}^n$ can be represented as

$$x = x_p + x_n$$

where $x_p \in R(A^T)$ which is

orthogonal projection
of x onto $R(A^T)$

$x_n \in N(A)$, which is

orthogonal projection
of x onto $N(A)$.

NL/12/11/25

The complete set of solⁿ of $Ax = b$
for A full row rank:

$$S = \{x \mid x = x_p + x_n\}$$

when x_p : particular solⁿ of $Ax = b$
 x_n : solution of homogeneous system $Ax = 0$

$$\begin{cases} Ax_p = b \\ Ax_n = 0 \end{cases} \quad \begin{matrix} Ax = b \\ A(x_p + x_n) = b \end{matrix}$$

The orthogonal projector P
that projects x onto $\mathcal{R}(A^T)$

$\uparrow$
 $Ax_n = 0$
 $\Downarrow$
 $x_n \in \mathcal{N}(A)$

$$\downarrow$$
$$\boxed{P = A^+ A}$$

when

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

← For full row rank matrix

$$\begin{aligned} x_p &= P x \\ &= A^+ A x \\ &= A^+ b \end{aligned}$$

$Ax = b$
← Particular solⁿ
 $Ax_p = b$

$$\begin{aligned} A A^+ b &= (A A^T) (A A^T)^{-1} b \\ &= b \end{aligned}$$

$$A(x_p + x_n) = 0 \quad \text{when} \quad x_n \in \mathcal{N}(A)$$

Any choice of $x_n \in \mathcal{N}(A)$ is a solⁿ

$$\downarrow \quad \text{of } Ax = b$$

We need a complete characterization
of x_n

• The complementary projection of P

$$\text{is : } I - P$$

• For any $x \in \mathbb{R}^n$

$$\boxed{(I - P)x = x_n} \leftarrow \text{complete characterization of all } x_n \in \mathcal{N}(A)$$

$$\downarrow$$

$$(I - P)x \in \mathcal{N}(A)$$

$$\begin{aligned} (I - A^+A)x &\rightarrow A(I - A^+A)x \\ &= (A - \underbrace{AA^+A}_{=A})x \\ &= 0 \end{aligned}$$

$$\Rightarrow (I - P)x \in \mathcal{N}(A)$$

$\rightarrow$ The solⁿ of $Ax = b$

$$\downarrow$$

$$A(x_p + x_n) = b$$

$$x = x_p + x_n$$

$$= A^+ b + (I - A^+ A) x \quad \text{for all } x \in \mathbb{R}^n.$$

↑
 Characterization of all possible
 solⁿ to $Ax = b$ where
 A is full row-rank.

constrained
 optimizer
 problem

$$\min \|x\|$$

→ Norm will be

$$\text{s.t. } Ax = b$$

minimum when $\|x\| = 0$

- Since the freedom is in choosing $x \in \mathbb{R}^n$ to characterize x , for min. norm

solⁿ let us choose $x = 0$

⇓

$$\bar{x} = A^+ b$$

Optimal
 value.

- Sometimes the constrained optimizer problem can be transformed to an unconstrained optimizer using "Lagrange Duality Theory".

$$\begin{cases} \min x^T x \\ \text{s.t. } Ax = b \Rightarrow Ax - b = 0 \end{cases}$$

→ converted to an unconstrained Optimizer
 using Lagrange variables.

The Lagrange funⁿ

$$\mathcal{L}(x, \lambda) := x^T x + \lambda^T (Ax - b)$$

new
variable

For unconstrained optimization

$$\nabla_x \mathcal{L}(\bar{x}, \bar{\lambda}) = 2\bar{x} + A^T \bar{\lambda} = 0$$

$$\nabla_\lambda \mathcal{L}(\bar{x}, \bar{\lambda}) = A\bar{x} - b = 0$$

$$\bar{x} = -\frac{1}{2} A^T \bar{\lambda}$$

$$A\left(-\frac{1}{2} A^T \bar{\lambda}\right) - b = 0$$

$$\Rightarrow -\frac{1}{2} A A^T \bar{\lambda} - b = 0$$

$$\Rightarrow \bar{\lambda} = -2 (A A^T)^{-1} b$$

$$\begin{aligned} \bar{x} &= A^T (A A^T)^{-1} b \\ &= A^+ b \end{aligned}$$

Least Square Problem

$$\min_{x \in \mathbb{R}} \|b - Ax\|_2^2$$

when A is full column rank.

• The solⁿ

$$\bar{x} = A^+ b$$

when $A^+ = (A^T A)^{-1} A^T$

• $A\bar{x}$ is the orthogonal projection of output data (b) onto image space of A
 $\mathcal{R}(A)$

Minimum Norm Problem

$$\min x^T x$$

$$\text{s.t. } Ax = b$$

when A is full row rank

• The solⁿ

$$\bar{x} = A^+ b$$

when $A^+ = A^T (A A^T)^{-1}$

• $\bar{x}$ is the orthogonal projection of a solution x onto the row space of $A = \mathcal{R}(A^T)$.

27th & 28th project? present?

