Summary – all the way to pseudo-inverse

Linear Equations

- Case 1: No. of equations = No. of variables
- Case 2: No. of equations > No. of variables
- Case 3: No. of equations < No. of variables

Case 1

Case 1: No. of equations = No. of variables

$m = n$

square matrix

A $m \times n$

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

If rank(A) = min(m,n) fullrank

$$(m = n)$$

$$\underset{(m\times n)}{\mathbf{A}}\ \underset{(n\times 1)}{\boldsymbol{x}} = \underset{(n\times 1)}{\mathbf{b}}$$

- **A** is full rank, that is $|A| \neq 0$, $\boldsymbol{x} = \mathbf{A}^{-1}\mathbf{b}$, unique solution

  rank = (m, ~~or n~~)

  $Ax = b$ Gauss elimination

- The case of A not full rank (rank < m).
  - Consistent $\rightarrow$ similar to less equations, more variables ∞ solutions
  - Inconsistent $\rightarrow$ no solution

Case 2

<u>Case 2:</u> No. of equations > No. of variables

$$x + y = 1$$
$$2x + 3y = 1$$
$$x + 2y = 1$$

$(m > n)$

$$\underset{(m\times n)}{\mathbf{A}}\ \underset{(n\times 1)}{\boldsymbol{x}} = \underset{(n\times 1)}{\mathbf{b}}$$

- Problem of not enough
- Not all equations can be satisfied
- Can be viewed as a no solution case
- Some form of optimization needed

$$x\begin{bmatrix}1\\2\\1\end{bmatrix} + y\begin{bmatrix}1\\3\\2\end{bmatrix} = \begin{bmatrix}1\\1\\1\end{bmatrix}$$

project b on the columns of A

Linear equtions

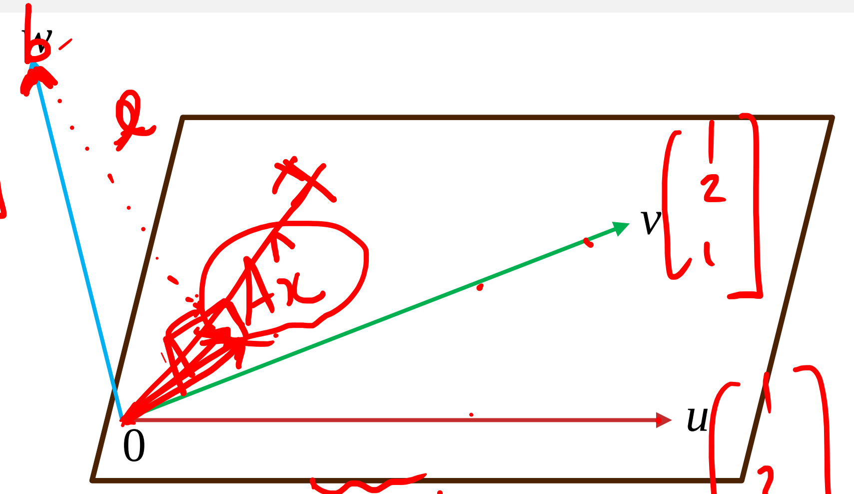
$e = b - A\hat{x}$

$e \perp\text{ar to } \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$

$e \perp\text{ar to } \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$

$\hat{x} = \begin{bmatrix} 0.667 \\ \\ \end{bmatrix} \quad \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

$\hat{y} = 0$



linear combination of colum of A.

column space of A

$$\underbrace{e = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \hat{x}\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} - \hat{y}\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}}$$

$\perp\text{ar} \quad \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \quad \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$

$e^T \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = 0 \qquad e^T \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = 0$

$$e^T \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = 0$$

$$e^T \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = 0$$

2 equations, 2 unknowns!

error ⊥ model

$$e = \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}_{\text{data}}^{\overbrace{}^{b}} - \overbrace{\hat{x} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} - \hat{y} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}}^{\text{model}}$$

$$e = \begin{bmatrix} 1 - \hat{x} - \hat{y} \\ 1 - 2\hat{x} - 3\hat{y} \\ 1 - \hat{x} - 2\hat{y} \end{bmatrix}$$

Case 2 continued – Optimization view

square of length of e

$$\text{Min } \underbrace{(\mathbf{A}\boldsymbol{x}\text{-}\mathbf{b})^T(\mathbf{A}\boldsymbol{x}\text{-}\mathbf{b})}$$

$$\text{Min } [(\boldsymbol{x}^T\mathbf{A}^T\text{-}\mathbf{b}^T)(\mathbf{A}\boldsymbol{x}\text{-}\mathbf{b}) = f(\mathbf{b})]$$

$$\text{Min } [(\boldsymbol{x}^T\mathbf{A}^T\mathbf{A}\boldsymbol{x}\text{-}2\mathbf{b}^T\mathbf{A}\boldsymbol{x}+\mathbf{b}^T\mathbf{b} = f(\mathbf{b})]$$

$$\nabla f = 0$$

$$2(\mathbf{A}^T\mathbf{A})\boldsymbol{x} - 2\mathbf{A}^T\mathbf{b} = 0$$

$(\mathbf{A}^T\mathbf{A})\boldsymbol{x} = \mathbf{A}^T\mathbf{b}$ (Remember n is the smaller number)

$$\boxed{\boldsymbol{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}}$$

n×m m×n n×m m×1

e = Ax-b

minimize "e" (?) why?

Ideal Ax-b = 0

minimize square of length $|e|^2 = e^Te$

Ax = b : m > n

m×n · n×1 m×1

why minimize e^Te?

Data as close as possible to model

Case 2 continued – Projection view

- Project **b** onto column space of **A**

$$\mathbf{b}=\mathbf{A}x'$$

$e = b - Ax'$ ⊥ar columns of A

A: $m \times n$ $m > n$

b: $m \times 1$

x: $n \times 1$

- Use orthogonality

$$\mathbf{A}^{\mathrm{T}}(\mathbf{b}-\mathbf{A}x') = 0$$

$$\underbrace{\mathbf{A}^{\mathrm{T}}\mathbf{b}}_{m \times 1} = \underbrace{(\mathbf{A}^{\mathrm{T}}\mathbf{A})}_{n \times m \;\, m \times n}\, x' \quad n \times 1$$

$n \times m \;\, m \times 1$

$$x' = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{b} \;\rightarrow \text{same solution}$$

WARNING! Only use as a <u>mnemonic</u>.

$$\underbrace{A}_{m\times n}\ \underbrace{x}_{n\times 1} = \underset{m\times 1}{b}.$$

$m > n.$

multiply by A^T

$$\underbrace{A^T A}_{n\times n}\hat{x} = \underbrace{A^T b}_{n\times 1}$$

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

$(A^T A)$ has to be invertible

Case 2 continued – Linear combination view

$$\mathbf{A}\,x = \mathbf{b}$$

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}\,x = \mathbf{A}^{\mathrm{T}}\mathbf{b}$$

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

- Here only a limited number of linear combinations (as many as the variables) are retained to solve the problem

Case 3

<u>Case 3:</u> No. of equations < No. of variables

$(m \times n)$

$Ax = b$

$m < n$

eqns < unknowns.

- Problem of plenty
- Infinite solutions
- What is the rationale for choosing a single solution from this infinity of solutions?

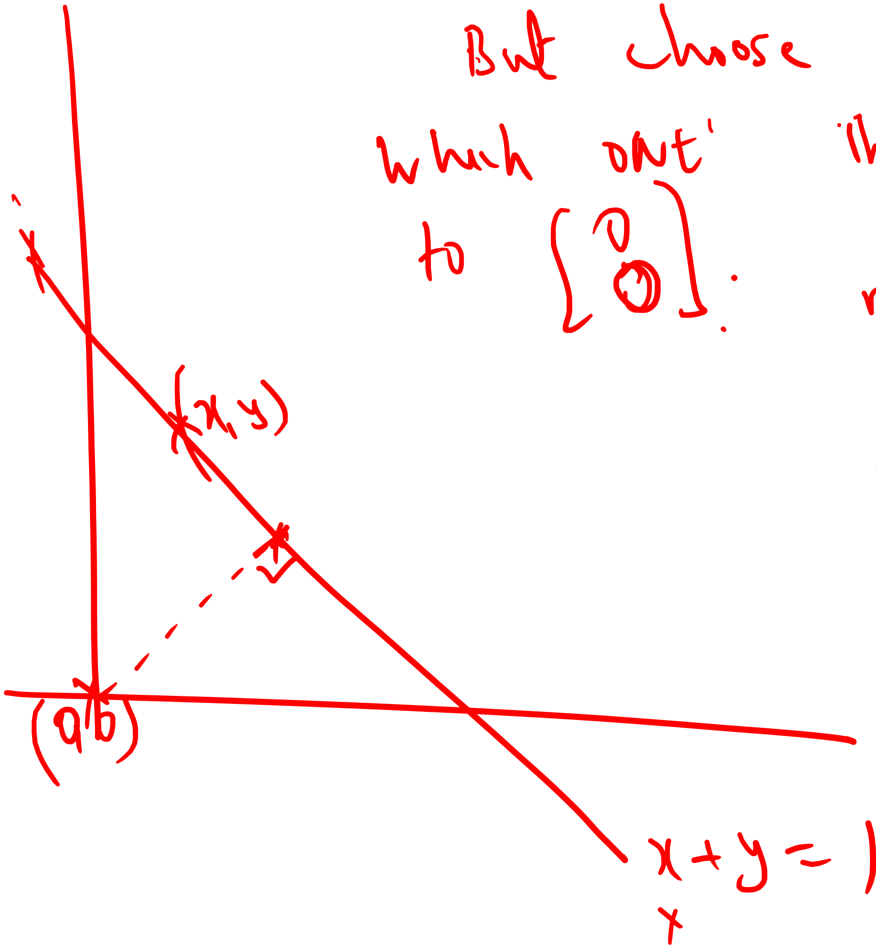
- choose x which has shortest length!

$x + y = 1$ Infinite solutions !

But choose "ONE"

which ONE' the one which is closest

to $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$.



$$\min \ (x-0)^2 + (y-0)^2 .$$

$$\text{s.t.} \quad x + y - 1 = 0$$

$$-\nabla f = \lambda \nabla g$$

$$x = 1/2$$

$$y = 1/2$$

Case 3 continued – Optimization view

- Pose an optimization problem
- Minimize ½ $\boldsymbol{x}^T\boldsymbol{x}$ -

  -such that $\mathbf{A}\boldsymbol{x}=\mathbf{b}$ $m \times n$

n variables

m constraints

m<n

- Min [$f(\mathbf{b})= ½\, \boldsymbol{x}^T\boldsymbol{x} + \lambda^T(\mathbf{A}\boldsymbol{x}-\mathbf{b})$]

$$\nabla f = 0$$

$$\boldsymbol{x} + \mathbf{A}^T\lambda = 0$$

$$\mathbf{A}\boldsymbol{x} = \mathbf{b}$$

Case 3 continued – Optimization view

$A: m \times n$

$AA^T: \underbrace{m \times m}$

$$\boldsymbol{x} = -\mathbf{A}^T \lambda$$

$$-\mathrm{AA}^T \lambda = \mathbf{b}$$

$$\lambda = -(\mathrm{AA}^T)^{-1}\mathbf{b}$$

$$\boxed{\boldsymbol{x} = -\mathbf{A}^T \lambda = \mathbf{A}^T\underbrace{(\mathrm{AA}^T)}_{m \times m}{}^{-1}\mathbf{b}}$$

- Again the inverse is of the smaller square matrix

$$\lambda = \underset{n \times m}{A^T} \underbrace{(\underset{m \times n}{A} \underset{n \times m}{A^T})}^{-1} \underset{m \times 1}{b} .$$

Case 3 continued – Principle of parsimony

$\boldsymbol{x}$ = Row space + Null space component

$= \boldsymbol{x}_\mathrm{r} + \boldsymbol{x}_\mathrm{N}$

Null space component implies $\mathbf{A}\boldsymbol{x}_\mathrm{N}=0$

$$\mathbf{A}\boldsymbol{x} = \mathbf{b}$$

$$\mathbf{A}(\boldsymbol{x}_\mathrm{r}+\boldsymbol{x}_\mathrm{N}) = \mathbf{b}$$

$$\mathbf{A}\boldsymbol{x}_\mathrm{r} = \mathbf{b}$$

Case 3 continued – Principle of parsimony

$\boldsymbol{x}_r$ is in row space

$$\boldsymbol{x}_r = \mathbf{A}^T y$$

$$\mathbf{A}\,\boldsymbol{x}_r = \mathbf{b}$$

$$\mathbf{A}\mathbf{A}^T y = \mathbf{b}$$

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

$$\boldsymbol{x}_r = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b} \rightarrow \text{same as previous result}$$

$\boldsymbol{x}_r$ is a unique solution

Generalization

- Can I combine all these results into one elegant result?

- Yes – notion of pseudo-inverse

- $\mathbf{A}x = \mathbf{b}$; $x = \mathbf{A}^{\dagger}\,\mathbf{b}$

$A^{\dagger}$ = pseudo-inverse

Pseudo-inverse

$$\underset{(m\times n)}{\mathbf{A}} = \underset{(m\times m)}{Q_1}\ \underset{(m\times n)}{\Sigma}\ \underset{(n\times n)}{Q^T_2}$$

$$\underset{}{\mathbf{A}^\dagger} = \underset{(n\times n)}{Q_2}\ \underset{(n\times m)}{\Sigma^\dagger}\ \underset{(m\times m)}{Q^T_1}$$

SVD of A.

+ pseudo inverse symbol (dagger)

m > n	m < n
$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ 0 & 0 & 0 \end{bmatrix}$	$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 \end{bmatrix}$ 3×4
$\Sigma^\dagger = \begin{bmatrix} 1/\sigma_1 & 0 & 0 & 0 \\ 0 & 1/\sigma_2 & 0 & 0 \\ 0 & 0 & 1/\sigma_3 & 0 \end{bmatrix}$	$\Sigma^\dagger = \begin{bmatrix} 1/\sigma_1 & 0 & 0 \\ 0 & 1/\sigma_2 & 0 \\ 0 & 0 & 1/\sigma_3 \\ 0 & 0 & 0 \end{bmatrix}$ 4×3.

$$\underset{(m\times n)}{\mathbf{A}}\ \underset{(n\times 1)}{\boldsymbol{x}} = \underset{(m\times 1)}{\mathbf{b}}$$

$$\boldsymbol{x} = \underset{(n\times m)}{\mathbf{A}^\dagger}\ \underset{(m\times 1)}{\mathbf{b}}$$