Lecture # 20

02/27/2023

We impose extra condition to extract unique solution vector $\underline{x}$ for a wide linear system:

$$\left.\begin{aligned} &\underset{\underline{x}\in\mathbb{R}^n}{\text{minimize}} \quad \|\underline{x}\|_2 \\ &\text{subject to} \quad A\underline{x} = \underline{b} \end{aligned}\right\}$$

This problem is called the "least norm problem"

This problem has a unique minimizer $\underline{x} \in \mathbb{R}^n$ provided $\boxed{\text{rank}(A) = m}$

__Claim:__ Solution of the least norm problem is given by the minimizer:

$$\underline{x}_{\text{least norm}} = A^T(\underbrace{AA^T}_{})^{-1}\underline{b}$$

nonsingular/invertible because A has full row rank m

$$= A^{\dagger}\underline{b}$$

pseudo-inverse of wide A

right-inverse for wide A

Compare this with the least square solution:

$$\hat{\underline{x}} = A^{\dagger}\underline{b} = (A^TA)^{-1}A^T\underline{b}$$

left inverse for tall A

Proof / derivation of this claim without calculus:

Let $\underline{x} \in \mathbb{R}^n$ is a solution of the wide linear system: $A\underline{x} = \underline{b}$ such that $\underline{x} \neq \underline{x}_{\text{least norm}}$

where $\underline{x}_{\text{least norm}} := A^T (AA^T)^{-1} \underline{b}$

$$\therefore \quad A\underline{x} = \underline{b} = A\underline{x}_{\text{least norm}}$$

$$\Leftrightarrow \quad \underbrace{A}_{m \times n} \underbrace{(\underline{x} - \underline{x}_{\text{least norm}})}_{n \times 1} = \underline{0}$$

next pg.

Then,

$$\|\underline{x}\|_2^2 = \|\underline{x}_{\substack{\text{least}\\\text{norm}}} + (\underline{x} - \underline{x}_{\substack{\text{least}\\\text{norm}}})\|_2^2$$

$$= \|\underline{x}_{\substack{\text{least}\\\text{norm}}}\|_2^2 + \|\underline{x} - \underline{x}_{\substack{\text{least}\\\text{norm}}}\|_2^2$$

$$+ 2(\underline{x} - \underline{x}_{\substack{\text{least}\\\text{norm}}})^T \underline{x}_{\substack{\text{least}\\\text{norm}}}$$

Cross term $= 0$

why? next pg.

The cross-term:

$$2\left(\underline{x} - \underline{x}_{\substack{\text{least}\\\text{norm}}}\right)^T \underline{x}_{\substack{\text{least}\\\text{norm}}}$$

$$= 2\left(\underline{x} - \underline{x}_{\substack{\text{least}\\\text{norm}}}\right)^T A^T (AA^T)^{-1} \underline{b}$$

$$= 2\left(\underbrace{A\underline{x} - A\underline{x}_{\substack{\text{least}\\\text{norm}}}}_{=\underline{0}}\right)^T (AA^T)^{-1} \underline{b}$$

$$= \underline{0}$$

$$\therefore \|\underline{x}\|_2^2 > \|\underline{x}_{\substack{\text{least}\\\text{norm}}}\|_2^2 \quad \text{since } \underline{x} \neq \underline{x}_{\substack{\text{least}\\\text{norm}}}$$

(Proved.)

$\therefore \quad \underline{x}_{\substack{\text{least} \\ \text{norm}}} = A^T (A A^T)^{-1} \underline{b}$ has the smallest possible 2-norm among all possible solutions of the wide linear system $A\underline{x} = \underline{b}$

Computing $\underline{x}_{\substack{\text{least} \\ \text{norm}}}$ using QR decomposition:

- Find QR decomposition of A^T: $\quad A^T = QR$

  with $Q \in \mathbb{R}^{n \times m}$ and $Q^T Q = I_{m \times m}$

  and $R \in \mathbb{R}^{m \times m}$, upper triangular non-singular

Then, $\underline{x}_{\substack{\text{least}\\\text{norm}}} = A^T (A A^T)^{-1} \underline{b}$

$= Q R^{-T} \underline{b}$

and $\| \underline{x}_{\substack{\text{least}\\\text{norm}}} \|_2 = \| Q R^{-T} \underline{b} \|_2 .$

$$= \sqrt{\underline{b}^T (R^{-T})^T \underbrace{Q^T Q}_{I} R^{-T} \underline{b}}$$

$$= \sqrt{\underline{b}^T R^{-1} R^{-T} \underline{b}}$$

$$= \| R^{-T} \underline{b} \|_2 .$$

- Ridge Regression / Tikhonov regularization / Regularized least squares :

Suppose $A \in \mathbb{R}^{m \times n}$ is wide, $m < n$, has full row rank m.

Define $J_1 := \| A\underline{x} - \underline{b} \|_2^2$

$J_2 := \| \underline{x} \|_2^2$

Least norm problem in prev. slides $\Longleftrightarrow$ $\underset{\underline{x} \in \mathbb{R}^n}{\text{minimize}} \quad J_2$

subject to $\quad J_1 = 0$

Regularized least squares problem :

$$\underset{\underline{x} \in \mathbb{R}^n}{\text{minimize}} \left\{ J_1 + \beta J_2 \right\}, \quad \beta > 0$$

regularization parameter

$$= \underset{\underline{x} \in \mathbb{R}^n}{\text{minimize}} \left\{ \underbrace{\| A\underline{x} - \underline{b} \|_2^2}_{\text{data fidelity}} + \beta \underbrace{\| \underline{x} \|_2^2}_{\text{regularization}} \right\}$$

Solution of the regularized least squares problem :

$$\underline{x}_\beta = \left(A^T A + \beta I \right)^{-1} A^T \underline{b} \quad \text{-----} \quad (*)$$

$$= A^T \left(A A^T + \beta I \right)^{-1} \underline{b} \quad \text{-----} \quad (**)$$

next pg.

Why?

Just re-write regularized least squares problem as standard/ordinary least squares:

$$\|A\underline{x} - \underline{b}\|_2^2 + \beta\|\underline{x}\|_2^2$$

$$= \Big\| \underbrace{\begin{bmatrix} A \\ \sqrt{\beta}\, I \end{bmatrix}}_{A_{new}} \underline{x} - \underbrace{\begin{pmatrix} \underline{b} \\ \underline{0} \end{pmatrix}}_{\underline{b}_{new}} \Big\|_2^2 \left\{ = \left\| \begin{pmatrix} A\underline{x} - \underline{b} \\ \sqrt{\beta}\,\underline{x} \end{pmatrix} \right\|_2^2 = \|A\underline{x} - \underline{b}\|_2^2 + \beta\|\underline{x}\|_2^2 \right.$$

$$= \|A_{new}\,\underline{x} - \underline{b}_{new}\|_2^2$$

$\therefore$ Standard / ordinary least squares solution:

$$\underline{x}_\beta = \left(A_{new}^T A_{new}\right)^{-1} A_{new}^T \underline{b}_{new}$$

$$= \left(\begin{bmatrix} A^T & \sqrt{\beta} I \end{bmatrix} \begin{bmatrix} A \\ \sqrt{\beta} I \end{bmatrix} \right)^{-1} \begin{bmatrix} A^T & \sqrt{\beta} I \end{bmatrix} \begin{pmatrix} \underline{b} \\ \underline{0} \end{pmatrix}$$

$$= \boxed{\left(A^T A + \beta I\right)^{-1} A^T \underline{b}} \quad \text{- - - - - -} \quad (*)$$

Why $(*) = (**)$?

$$\underbrace{\beta M + MNM}$$

one way

another way

$$M(\beta I + NM) = (\beta I + MN) M$$

$$\Rightarrow \underbrace{(\beta I + NM)^{-1} M = M (\beta I + NM)^{-1}}$$

Now substitute : $M = A^T$, $N = A$

to obtain $(*) = (**)$

$$\underline{x}_\beta = \left(A^T A + \beta I\right)^{-1} A^T \underline{b} \quad \text{-----} \quad (*)$$

$$= A^T \left(A A^T + \beta I\right)^{-1} \underline{b} \quad \text{-----} \quad (**)$$

When $\beta \downarrow 0$, then $\underline{x}_\beta \longrightarrow A^+ \underline{b}$

appropriate pseudo-inverse depending on whether A is tall or wide

Summary of Solving Linear Systems

Solving linear system : $\underbrace{A}_{m\times n}\,\underbrace{\underline{x}}_{n\times 1} = \underbrace{\underline{b}}_{m\times 1}$

Determined $(m = n)$

If $\det(A) \neq 0$,

then unique solution $\underline{x}$

$$\underbrace{\underbrace{A}_{n\times n}\;\underbrace{\underline{x}}_{n\times 1} = \underbrace{\underline{b}}_{n\times 1}}_{\text{LU decomposition, Gauss elimination}}$$

Over-determined $(m > n)$:

$$\min_{\underline{x} \in \mathbb{R}^n} \| A\underline{x} - \underline{b} \|_2^2$$

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

$$= A^{+}\, \underline{b}$$

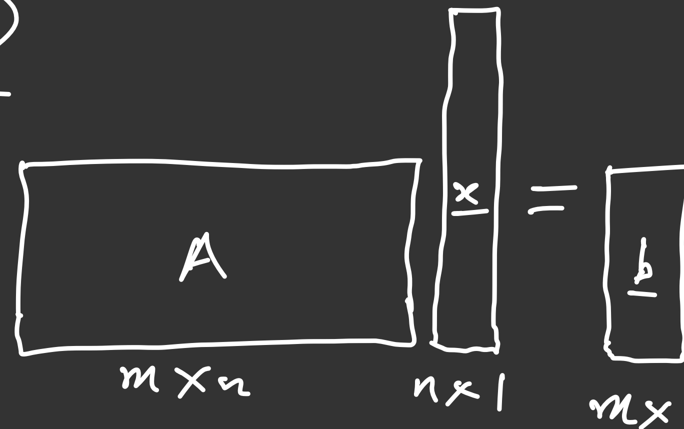
$$\underbrace{\underset{m \times n}{A}\ \underset{n \times 1}{\underline{x}} \approx \underset{m \times 1}{\underline{b}}}_{\text{Least squares.}}$$

next pg.

Underdetermined ($m < n$)

$$\min_{\underline{x} \in \mathbb{R}^n} \|\underline{x}\|_2$$

such that $A\underline{x} = \underline{b}$



$$\underline{x}_{\text{least norm}} = \underbrace{A^T(AA^T)^{-1}}\,\underline{b}$$

$$= A^+\,\underline{b}$$

next pg.

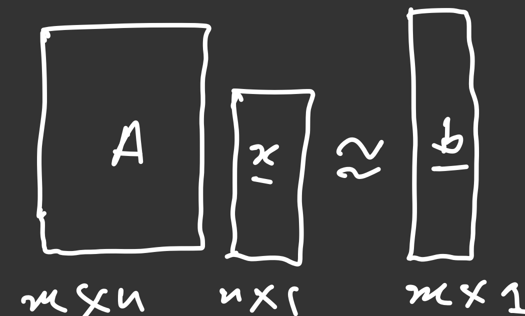
Noisy measurements:

$$\min_{\underline{x}\in\mathbb{R}^n} \|A\underline{x}-\underline{b}\|_2^2 + \beta\|\underline{x}\|_2^2$$

Solution:

$$\underline{x}_\beta = \underbrace{(A^TA+\beta I)^{-1}A^T\underline{b}} \xrightarrow{\beta\downarrow 0} A^+\underline{b}$$

either



or