Lecture 8

Least-norm solutions of undetermined Equations

★ Underdetermined linear equations

⇒ We consider

$$y = Ax$$

where $A \in \mathbb{R}^{m \times n}$ is fat $(m < n)$

→ there are more variables than equations.

→ x is underspecified (i.e. many choices of x lead to the same y)

⇒ We'll assume that A is full rank (m), so for each $y \in \mathbb{R}^m$, there is a solution set of all solutions has form

$$\{x \mid Ax = y\} = \{x_p + z \mid z \in N(A)\}$$

where x_p is any ('particular') solution (i.e. $Ax_p = y$)

⇒ Solution has $\dim N(A) = n - m$ ⇒ degree of freedom

→ Can choose z to satisfy other specs or optimize among solutions.

★ Least-norm Solution

⇒ One particular solution is

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

$\{AA^T$ is invertible since A is full rank$\}$

⇒ x_{ln} is solution of optimization problem

$$\boxed{\begin{array}{l} \text{minimize } \|x\| \\ \text{subject to } Ax = y \end{array}}$$ where $x \in R^n$

Proof

⇒ Suppose $Ax = y$, so $A(x - x_{ln}) = 0$,

$$\Rightarrow (x - x_{ln})^T x_{ln} = (x - x_{ln})^T A^T (AA^T)^{-1} y$$

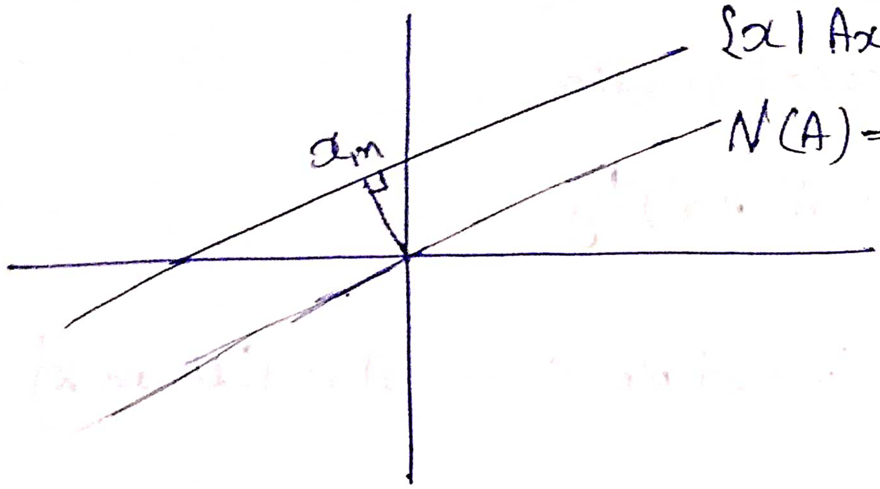
$$= (A(x - x_{ln}))^T (AA^T)^{-1} y$$

$$= 0$$

$\rightarrow (x - x_{ln}) \perp x_{ln}$

⇒ So, $\|x\|^2 = \|x_{ln} + x - x_{ln}\|^2 = \|x_{ln}\|^2 + \|x - x_{ln}\|^2 \geq \|x_{ln}\|^2$

$\|x_{ln}\|^2 \leq \|x\|$ $\{x_{ln}$ has smallest norm of any solution$\}$



- <u>Orthogonality Condition</u>: $x_{ln} \perp N(A)$

- <u>Projection interpretation</u>: x_{ln} is projection of 0 on solution set $\{x \mid Ax = y\}$

$\Rightarrow$ $A^{+} = A^T(AA^T)^{-1}$ is called the pseudo-inverse of full rank, fat A.

$\Rightarrow$ $A^T(AA^T)^{-1}$ is a right inverse of A.

$\Rightarrow$ $I - A^T(AA^T)^{-1}A$ gives projection onto $N(A)$.

★ <u>Least-norm solution via QR factorization</u>

$\Rightarrow$ find QR factorization of A^T i.e. $A^T = QR$ with

$\to Q \in R^{n \times m}$, $Q^TQ = I_m$

$\to R \in R^{m \times m}$, upper triangular, non singular

$\Rightarrow$ then

$x_{ln} = A^T(AA^T)^{-1}y = QR^{-T}y$

$\|x_{ln}\| = \|R^{-T}y\|$

★ Derivation via Lagrange multipliers

⇒ least-norm solution solves optimization problem

$$\text{minimize } x^T x$$
$$\text{subject to } Ax = y$$

⇒ Introduce Lagrange multipliers:

$$L(x, \lambda) = x^T x + \lambda^T (Ax - y)$$

→ Objective

→ Lagrange multiplier

→ Constraints

⇒ optimality conditions are.

$$\nabla_x L = 2x + A^T \lambda = 0 \Rightarrow x = -A^T \lambda / 2$$

$$\nabla_\lambda L = Ax - y = 0$$

$$\lambda = -2(AA^T)^{-1} y$$

⇒ hence $\boxed{x = A^T (AA^T)^{-1} y}$

$\|x\|_\infty = \max_i |x_i|$ {Infinity norm)

$\|x\|_1 = \sum_i |x_i|$ {Manhattan norm}

$\|x\|_2 = \sum_i |x_i|^2$ {Euclidean norm}

★ Relation to regularized least-squares

⇒ Suppose $A \in \mathbb{R}^{m \times n}$ is fat, full rank.

⇒ define $J_1 = \|Ax - y\|^2$; $J_2 = \|x\|^2$

⇒ least-norm solution minimizes J_2 with $J_1 = 0$

⇒ minimizer of weighted-sum objective

$$J_1 + \mu J_2 = \|Ax - y\|^2 + \mu \|x\|^2 \text{ is}$$

$$x_\mu = (A^T A + \mu I)^{-1} A^T y$$

⇒ fact: $x_\mu \to x_{ln}$ as $\mu \to 0$

{regularized solution converges to least-norm solution as $\mu \to 0$}

⇒ In matrix terms as $\mu \to 0$

$$(A^T A + \mu I)^{-1} A^T \to A^T (A A^T)^{-1}$$

{for full rank, fat A}

★ General norm minimization with equality constraints

⇒ Consider problem,

$$\text{minimize } \|Ax-b\|$$
$$\text{subject to } Cx = d$$

with variable x

⇒ Includes least-square and least-norm problems as special case.

⇒ Equivalent to

$$\text{minimize } \tfrac{1}{2}\|Ax-b\|^2$$
$$\text{subject to } Cx = d$$

⇒ Lagrangian is

$$L(x,\lambda) = \tfrac{1}{2}\|Ax-b\|^2 + \lambda^T(Cx-d)$$

$$= \tfrac{1}{2}x^TA^TAx - b^TAx + \tfrac{1}{2}b^Tb + \lambda^TCx - \lambda^Td$$

$$\nabla_x L = A^TAx - A^Tb + C^T\lambda = 0 \quad \Big| \quad \nabla_\lambda L = Cx - d = 0$$

⇒ Write in block matrix form as:

$$\begin{bmatrix} A^TA & C^T \\ C & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} A^Tb \\ d \end{bmatrix}$$

$$\begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} A^TA & C^T \\ C & 0 \end{bmatrix}^{-1} \begin{bmatrix} A^Tb \\ d \end{bmatrix} \quad \{\text{If invertible}\}$$

⇒ If A^TA is invertible, we can derive a more explicit (and complicated) formula for x

$$x = (A^TA)^{-1}(A^Tb - C^T\lambda)$$

⇒ Substitute into $Cx = d$ we get

$$C(A^TA)^{-1}(A^Tb - C^T\lambda) = d$$

$$\Rightarrow \lambda = \left(C(A^TA)^{-1}C^T\right)^{-1}\left(C(A^TA)^{-1}A^Tb - d\right)$$

⇒ recover x from equation above:

$$\boxed{x = (A^TA)^{-1}\left(A^Tb - C^T\left(C(A^TA)^{-1}C^T\right)^{-1}\left(C(A^TA)^{-1}A^Tb - d\right)\right)}$$