10. Surveys of Difficulties with $Ax=b$

Our main agenda is to solve $Ax=b$. Ordinary elimination might not compute x - There might be too many equations, no solutions, & so on.

In deep learning, we have too many solutions - and we want one to generalize well unseen test data.

~~The pseudoinverse of A never per is a good solution~~

0. $x = A^+ b$ [The pseudoinverse way]

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

1. condition number $= \dfrac{\sigma_1}{\sigma_n}$

If this is not too large (<1000) $\rightarrow$ elimination succeeds with a good accuracy.

2. $m > n = r$ $\quad A = [\quad]$

- Too many equations
- Eg. least squares

$\rightarrow$ If columns are independent ($n=r$) and not not too ill conditions,

$$A^TA\hat{x} = A^Tb$$
$$\hat{x} = (A^TA)^{-1}A^Tb$$

$\rightarrow$ We project $\vec{b}$ onto column space of A and find the nearest solution

3. $m < n$

- Short and wide $A = [\qquad\qquad]$

- too many solutions, if it has one.
- choose best x for our purpose

$x = x^T = A^+ b$ = has minimum l^2 norm solution

$x = x_1$ = minimum l^1 norm solution

4. The columns are in bad condition $\frac{\sigma_1}{\sigma_r} >>$

→ Orthogonalize the columns using Gram-Schmidt / Householder algorithm

$A = QR$

5. Near singular / Inverse problems / ADD penalty

• Inverse problem : output known; find the system

minimize $||Ax - b||^2 + \underbrace{\delta^2 ||x||^2}_{\text{Penalty term}}$ ⇒ more well conditioned

6. A is too big ⇒ KRYLOV [iterative method]

→ ~~random sampling~~

7. Way too big → [Random numerical algebra]

Regularization:

minimize $\left(\|Ax-b\|^2 + \delta^2 \|x\|_2^2\right)$ choose $\delta > 0$.

$$\begin{bmatrix} A \\ \delta I \end{bmatrix} [x] = \begin{bmatrix} b \\ 0 \end{bmatrix}$$

$$A^* x = b^*$$

minimize $\|Ax-b\|^2 + \delta^2 \|x\|^2$ — ①

let,
A is 1×1 matrix, $A = [\sigma]$

$$f = \sigma^2 x^2 + \delta^2 x^2 - 2\sigma b x + b^2$$
$$f' = 2\sigma^2 x + 2\delta^2 x - 2\sigma b = 0$$
$$\Rightarrow (\sigma^2 + \delta^2)\, x = \sigma b$$

$$(\sigma^2 + \delta^2) x_\delta = \sigma b$$
$$x_\delta = \left(\frac{\sigma}{\sigma^2 + \delta^2}\right) b.$$

if $\sigma \neq 0$
$$x = \frac{1}{\sigma} b$$
if $\sigma = 0$
$$x = 0.$$
this is the pseudo inverse.

Staying in l^2 norm:

$$(A^T A + \delta^2)^{-1} A^T \to A^+ \text{ as } \delta \to 0$$

minimize $\|Ax-b\|_2^2 + \sigma^2 \|x\|_1^2 \implies$ LASSO

Conclusion,
$$\lim_{\delta \to 0} (A^T A + \delta^2 I)^{-1} A^T = A^+$$

Scanned with CamScanner