Oct 7, 2020

LLS: $\hat{x} = \underset{x}{\operatorname{argmin}} \|Ax-b\|_2^2$ A $m \times n$ $m > n$

A full rank

$\hat{x} = A^+ b$ $\hat{x} = R^{-1}Q^T b$ $r = b - A\hat{x} = (I - QQ^T)b$

① conditioning: how sensitive problem is to perturbations

② stability: how accurate our algorithm is

backward stable alg $\Rightarrow$ relative error is $O(\kappa \cdot \varepsilon_{mach})$

LLS: $(A, b) \to (\hat{x}, A\hat{x}) = (A^+b, AA^+b)$

perturb A or b, how does $\hat{x}$ change? ($A\hat{x}$ change?)

$m = n \Rightarrow$ should get $\kappa_2(A) = \|A\|_2 \|A^{-1}\|_2 = \sigma_1/\sigma_n$

Ex: $A = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ $b = \begin{pmatrix} \varepsilon \\ 1 \end{pmatrix}$ $A^TA = [1]$ $A^Tb = \varepsilon$

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

$b + \delta b = \begin{pmatrix} 2\varepsilon \\ 1 \end{pmatrix} \Rightarrow$ $\tilde{x} = 2\varepsilon$ $\|\tilde{x} - \hat{x}\| / \|\hat{x}\| = 1$

$\|\delta b\| / \|b\| = O(\varepsilon)$

Problem! Projection of b onto range(A) is small ($O(\varepsilon)$) relative to the size of b

$A\hat{x} = \begin{pmatrix} \varepsilon \\ 0 \end{pmatrix} \qquad A\tilde{x} = \begin{pmatrix} 2\varepsilon \\ 0 \end{pmatrix} \qquad \frac{\|A\tilde{x} - A\hat{x}\|}{\|A\hat{x}\|} = O(1)$

$\hat{x} = R^{-1}Q^T b$

$k_2(A) = \sigma_1 / \sigma_n \qquad \boxed{A} = \boxed{U}\ \underset{\Sigma}{\boxed{\diagdown}}\ \boxed{V^T}$

if $k_2(A)$ large, A square, except problems

$\|A\|_2 = \|QR\|_2 = \|R\|_2 = \sigma_1$

$\|A^+\|_2 = \|R^{-1}Q^T\|_2 = \|QR^{-T}\|_2 = \|R^{-T}\|_2 = \|R^{-1}\|_2 = 1/\sigma_n$

$k_2(R) = \sigma_1/\sigma_n = k_2(A)$

What is happening when $\sigma_n \ll \sigma_1$

A is nearly rank deficient

small perturbations to A can cause relatively large changes to range(A)

Lemma: Let $M = M^T$ with $\|M\|_2 < 1$. Then $(I-M)^{-1} = \sum_{i=0}^{\infty} M^i$

Proof: $M = V \Lambda V^T$ $\quad M^2 = V \Lambda V^T V \Lambda V^T = V \Lambda^2 V^T$ $\quad M^i = V \Lambda^i V^T$

$$S_n = \sum_{i=0}^{n} M^i = \sum_{i=0}^{n} V \Lambda^i V^T = V \left(\sum_{i=0}^{n} \Lambda^i \right) V^T$$

$$\sum_{i=0}^{n} (\Lambda^i)_{jj} = \frac{1 - \lambda_j^{n+1}}{1 - \lambda_j}$$

$\{\sigma_i\} = \{|\lambda_i|\}$ $\quad \sigma_1 < 1 \Rightarrow |\lambda_i| < 1 \Rightarrow S_n$ converges

$$(I-M) S_n = (I-M) \sum_{i=0}^{n} M^i = \sum_{i=0}^{n} M^i - \sum_{i=1}^{n+1} M^i = I - M^{n+1}$$

$$M^{n+1} = V \Lambda^{n+1} V^T \to 0 \Rightarrow S_n \to (I-M)^{-1}$$

Corollary: $\|(I-M)^{-1}\|_2 \leq \dfrac{1}{1 - \|M\|_2}$

Corollary: $(I-M)^{-1} = I + M + O(\|M\|_2^2)$

$\|\cdot\|$ $\quad \|AB\| \leq \|A\| \|B\|$

$$(A, b) \to (A^+ b,\ AA^+ b) = (\hat{x},\ A\hat{x})$$

perturbed input $(A+\delta A,\ b+\delta b) \to (\tilde{x},\ A\tilde{y})$

$$\tilde{x} = \left[(A+\delta A)^T(A+\delta A)\right]^{-1}(A+\delta A)^T(b+\delta b)$$

$$\approx \left[A^TA + \underbrace{\delta A^T A}_{B} + \underbrace{A^T\delta A}_{B^T} + \cancel{\delta A^T\delta A}\right]^{-1}\left(A^Tb + A^T\delta b + \delta A^T b + \cancel{\delta A^T\delta b}\right)$$

$$= \left[(A^TA)\left(I + (A^TA)^{-1}(B+B^T)\right)\right]^{-1}\left(A^Tb + A^T\delta b + \delta A^T b\right)$$

$$= \left(I + \underbrace{(A^TA)^{-1}(B+B)^T}_{M}\right)^{-1}(A^TA)^{-1}\left(A^Tb + A^T\delta b + \delta A^+ b\right)$$

$$\approx \left(I - M + \cancel{O(\|M\|_2^2)}\right)\left(\hat{x} + (A^TA)^{-1}(A^T\delta b + \delta A^T b)\right)$$

$$= \hat{x} - M\hat{x} + (A^TA)^{-1}\left(A^T\delta b + \delta A^T b\right) + \text{higher-order terms}$$

$$\tilde{x} - \hat{x} = -M\hat{x} + (A^TA)^{-1}\underbrace{\left(A^T\delta b + \delta A^T b\right)}_{\delta y}$$

$y = A^T b$

$\|\delta y\| / \|y\|$

$$\frac{\|\delta y\|}{\|y\|} = \frac{\|A^T\delta b + \delta A^T b\|}{\|A^T b\|} \le \frac{\|A\|\,\|b\|}{\|A^T b\|}\left(\underbrace{\frac{\|\delta A\|}{\|A\|}}_{\text{rel. change } A} + \underbrace{\frac{\|\delta b\|}{\|b\|}}_{\text{rel change } B}\right)$$

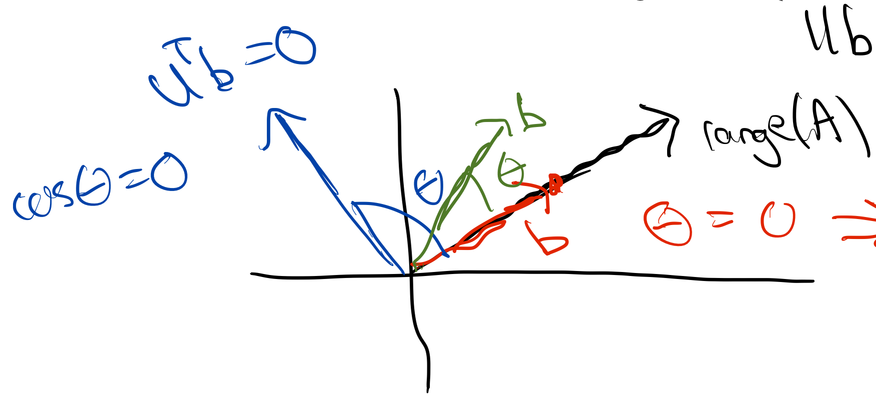
$$\frac{\|A\|\,\|b\|}{\|A^T b\|} = \frac{\sigma_1 \|b\|}{\|V\Sigma U^T b\|} = \frac{\sigma_1\|b\|}{\|\Sigma U^T b\|} \le \frac{\sigma_1 \|b\|}{\sigma_n \|U^T b\|}$$

∞

$$\frac{\|\tilde{x} - \hat{x}\|}{\|\hat{x}\|} \le \varepsilon\left[\frac{2\kappa_2(A)}{\cos\theta} + \tan\theta\,\kappa_2^2(A)\right] + O(\varepsilon^2)$$

$$\sin\theta = \frac{\|r\|_2}{\|b\|_2} \qquad \varepsilon = \max\left(\frac{\|\delta A\|}{\|A\|}, \frac{\|\delta b\|}{\|b\|}\right)$$

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



$\theta = 0 \Rightarrow \cos\theta = 1$

$\tan(\theta) = 0$

What can we do for sensitivity?

$b \perp \text{range}(A) \iff A^T b = 0 \iff U^T b = 0$

not much to do

if b near range(A) and $\kappa_2(A)$ large

accuracy: Householder