Oct 9, 2020

LLS: $\hat{x} = \arg\min_x \|Ax-b\|_2^2$ $A = QR$ $R\hat{x} = Q^T b$

Sensitivity:

① $\boxed{A \text{ nearly singular}}$ ② b not be close to range(A)

Want: backward stable

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

$A = QR$ — Is $\hat{Q}\hat{R} \approx A$? / Is $\hat{Q}^T\hat{Q} \approx I$?

CGS / MGS

Householder / Givens

Householder $Q_j = I - 2v_jv_j^T$

$x = a_j$ $v_j = \text{sign}(x_1)\|x\|_2 e_1 + x_1$ $v_j = v_j / \|v_j\|_2$

$\tilde{Q} = Q + E$, $\|E\| = O(\varepsilon)$

$fl(\tilde{Q}A) = fl((Q+E)A) = (Q+E)A + F = QA + EA + F$

$= Q(A + \underline{Q^T E A} + \underline{Q^T F})$ $(QQ^T = I)$

$\|Q^T EA\|_2 \le \|Q^T\| \|E\| \|A\| = O(\varepsilon \|A\|)$

$|q_i^T a_j| \le \|q_i\|_2 \|a_j\| = \|a_j\|_2$

$\|Q^T F\|_2 \le \|F\|_2$ $|F_{ij}| = |q_i^T a_j| |\delta|$ $|\delta| = O(\varepsilon)$ $\|F\|_2 \le O(\varepsilon \cdot \|A\|_2)$

$Q(A+G)$ $\quad \|G\|_2 = O(\varepsilon \cdot \|A\|_2)$ $\quad G = Q^T E A + Q^T F$

$fl(\tilde{Q}_2 \tilde{Q}_1 A) = fl(\tilde{Q}_2(Q_1 A + G_1))$ $\quad \|G_1\|_2 = O(\varepsilon \|A\|_2)$

$= Q_2(Q_1 A + G_1) + G_2$

$$\|G_2\|_2 = O(\varepsilon \cdot \|Q_1 A + G\|_2) = O(\varepsilon(\overset{1}{\cancel{\|Q_1\|}}\|A\| + \overset{O(\varepsilon)}{\cancel{\|G_1\|}}))$$

$= O(\varepsilon \|A\|_2) + O(\varepsilon^2)$

$$Q_2(Q_1 A + G_1) + G_2 = Q_2 Q_1 A + \underbrace{Q_2 G_1}_{O(\varepsilon \|A\|)} + G_2) \quad O(\varepsilon \|A\|)$$

Multiplying by seq. of orthog matrices is BWS

Cor: $fl(\tilde{Q}_1 \cdots \tilde{Q}_n e_j) = q_j + g$ $\quad \|g\| = O(\varepsilon \|A\|_2)$

$\tilde{Q}^T \tilde{Q} \approx I$

Ill-conditioning

A is $m \times n$

$\|Ax - b\|_2$ $\quad$ A. full rank $\Rightarrow \hat{x} = V \Sigma^{-1} U^T b$

A rank $r < n \Rightarrow A = $

$$
A = \begin{array}{c} \\ m \end{array}
\begin{array}{c} \begin{array}{ccc} r & n-r & m-n \end{array} \\ \begin{bmatrix} U_1 & U_2 & U_3 \end{bmatrix} \end{array}
\begin{array}{c} \\ r \\ n-r \\ m-n \end{array}
\begin{array}{c} \begin{array}{cc} r & n-r \end{array} \\ \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \\ \multicolumn{2}{c}{0} \end{bmatrix} \end{array}
\begin{array}{c} \\ n \end{array}
\begin{array}{c} \begin{array}{cc} r & n-r \end{array} \\ \begin{bmatrix} V_1^T & V_2^T \end{bmatrix} \end{array}
$$

$$\|U \Sigma V^T x - b\|_2 = \|\Sigma V^T x - U^T b\|_2$$

$$= \left\| \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \\ \multicolumn{2}{c}{\cancel{0}} \end{pmatrix} [V_1^T \; V_2^T] \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} U_1^T b \\ U_2^T b \\ \cancel{U_3^T b} \end{pmatrix} \right\|_2$$

$$= \left\| \begin{pmatrix} \Sigma_1 & 0 \\ \cancel{0} & \cancel{0} \end{pmatrix} [V_1^T x_1 \; V_2^T x_2] - \begin{pmatrix} U_1^T b \\ \cancel{U_2^T b} \end{pmatrix} \right\|_2$$

$$= \|\Sigma_1 V_1^T x_1 - U_1^T b\|_2 \qquad \hat{x}_1 = V_1 \Sigma_1^{-1} U_1^T b$$

$$\hat{x} = \begin{pmatrix} V_1 \Sigma_1^{-1} U_1^T b \\ \hat{x}_2 \end{pmatrix} \qquad \hat{x}_2 = V_2 y \qquad A \begin{pmatrix} 0 \\ \hat{x}_2 \end{pmatrix} = 0 \qquad V_2 \underset{r}{y} = \hat{x}_2$$

$$A \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix} = U \Sigma \begin{pmatrix} V_1^T \hat{x}_1 & V_2^T \hat{x}_2 \end{pmatrix}$$

V_2 basis for null(A)

$$\|\hat{x}\|_2^2 = \Big\| \Big(\underbrace{V_1 \Sigma_1^{-1} U_1^T b}_{\hat{x}_1} \; \hat{x}_2 \Big) \Big\|_2^2 = \|\hat{x}_1\|_2^2 + \|\hat{x}_2\|_2^2$$

$\hat{x}_2 = 0$

"Smallest" solution $\hat{x} = \begin{pmatrix} \hat{x}_1 \\ 0 \end{pmatrix}$

no component in null space

$A^+ = V \Sigma^+ U^T, \quad \Sigma^+_{ii} = \begin{cases} 1/\sigma_i & \sigma_i > 0 \\ 0 & \sigma_i = 0 \end{cases}$

What if A has full rank but $\sigma_1 / \sigma_n = \kappa_2(A)$ large?

One idea: drop small singular values "truncated SVD"

$$A^+_\tau = V \Sigma^+_\tau U^T \qquad \Sigma^+_\tau = \begin{cases} 1/\sigma_i & \sigma_i > \tau \\ 0 \end{cases}$$

$\hat{x} = A^+_\tau b \qquad \tilde{x} = A^+_\tau (b + \delta b)$ max $1/\sigma_i$ $\sigma_i \geq \tau$ $\leq 1/\tau$

$$\|\hat{x} - \tilde{x}\| = \|A^+_\tau (b - (b + \delta b))\|_2 \leq \|A^+_\tau\|_2 \|\delta b\|_2$$

$$\leq \|\delta b\|_2 / \tau$$

Tradeoff: larger $\tau \Rightarrow \|A\hat{x} - b\|_2$ larger

Principal components regression

$$A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix} \quad \text{suppose } \frac{1}{m}\sum_{i=1}^{m} a_i^T = 0$$

PCA: $A = U\Sigma V^T$

loadings: $V_k = [v_1 \cdots v_k]$ PCs: $U_k\Sigma_k = [u_1\sigma_1 \cdots u_k\sigma_k]$

PCR: $\min_x \|U_k\Sigma_k x - b\|_2^2$

$$\hat{x} = \Sigma_k^{-1}U_k^T b$$

$$A_\tau^+ b = V\Sigma_\tau^+ U^T b = V_k\Sigma_k^{-1}U_k^T b = V_k\hat{x}$$

$\tau = \sigma_k + \varepsilon$

$$\min_{x} \|Ax - b\|_2^2 + \underline{\lambda^2 \|x\|_2^2}$$

regularization

ℓ_2-reg

Tikhonov reg.

weight decay

$$\left\| \begin{pmatrix} A \\ \lambda I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|_2^2 = \left\| \begin{matrix} Ax - b \\ \lambda x \end{matrix} \right\|_2^2 = \|Ax - b\|_2^2 + \lambda^2 \|x\|_2^2$$