PSet 3

Tuesday, March 14, 2023 8:12 PM

1

a) $\|\Delta x\|_2 = \|A^{-1}\Delta b\|_2 = \frac{\|A^{-1}\Delta b\|_2 \|\Delta b\|_2}{\|\Delta b\|_2}$

from pset2 $\|A\|_2 = \max \frac{\|Ax\|_2}{\|x\|_2}$

$\Rightarrow \max \frac{\|A^{-1}\Delta b\|_2}{\|\Delta b\|_2} = \|A^{-1}\|_2 \Rightarrow$

$\Rightarrow \max \frac{\|A^{-1}\Delta b\|_2}{\|\Delta b\|_2}\|\Delta b\|_2 = \|A^{-1}\|_2\|\Delta b\|_2$

$\therefore \|\Delta x\|_2 \leq \|A^{-1}\|_2\|\Delta b\|_2$

since $\|A^{-1}\|_2\|\Delta b\|_2$ is the max

similarly

$\|b\|_2 = \|Ax\|_2 = \frac{\|Ax\|_2\|x\|_2}{\|x\|_2} \Rightarrow$

$\Rightarrow \max \frac{\|Ax\|_2\|x\|_2}{\|x\|_2} = \|A\|\,\|x\|_2$

$\therefore \|b\|_2 \leq \|A\|_2\|x\|_2$

since $\|A\|_2\|x\|_2$ is the max of $\|b\|_2$

b) From a) we know:

$\|\Delta x\|_2 \leq \|A^{-1}\|_2\|\Delta b\|_2$

$\|b\|_2 \leq \|A\|_2\|x\|_2 \Rightarrow \frac{1}{\|x\|_2} \leq \frac{\|A\|_2}{\|b\|_2}$

and if we multiply $\|\Delta x\|_2 * \frac{1}{\|x\|_2}$ we know that the product will also be $\leq \frac{\|A\|_2}{\|b\|_2} * \|A^{-1}\|_2\|\Delta b\|_2 = \|A\|_2\|A^{-1}\|\frac{\|\Delta b\|_2}{\|b\|_2}$

c) We know that the induced form of a has an ℓ_2 norm of σ_1. Since A is invertible, A^{-1} has single values of $\frac{1}{\sigma_m} \cdots \frac{1}{\sigma_1}$ so the ℓ_2 norm of $A^{-1} = \frac{1}{\sigma_m}$.

$\therefore \|A\|_2\|A^{-1}\|_2 = \frac{\sigma_1}{\sigma_m}$ where σ_1 is the largest singular value of A and σ_m is the smallest. Because A is invertible, it follows that $\|A^{-1}\|_2\|(A^T)^{-1}\|_2 = \frac{1}{\sigma_m} * \sigma_1 = \frac{\sigma_1}{\sigma_m} = K(A)$

d) To maximize, we want the above equations to be $=$ instead of $\leq$. This means we want $\|b\| = \frac{1}{\sigma_m}$ and $\|\Delta b\| = \frac{1}{\sigma_1}$ so $b = U\begin{pmatrix}0\\\vdots\\1\end{pmatrix}$ and $\Delta b = U\begin{pmatrix}1\\\vdots\\0\end{pmatrix}$

2 We know that $K(A)$ is the ratio of the largest to the smallest singular value of A.

If we plug in A^TA we get:

$$K(A^TA) = \frac{\sigma_{max}(A^TA)}{\sigma_{min}(A^TA)} = \frac{\sigma_{max}(A)^2}{\sigma_{min}(A)^2} = K(A)^2$$

If we then take $K(\hat{R}) = K(A)$ and square both sides we have $K(\hat{R})^2 = K(A)^2 \Rightarrow K(\hat{R})^2 = K(A^TA)$

If we substitute in $A = \hat{Q}\hat{R}$ we get:

$$K(R)^2 = K(\hat{R}^T\cancel{\hat{Q}^T\hat{Q}}\hat{R}) \Rightarrow K(R)^2 = K(\hat{R}^TR)$$

using the same proof as above, we get:

$$K(\hat{R}^T\hat{R}) = K(\hat{R})^2$$

and if we plug it back in, we get

$$K(\hat{R})^2 = K(\hat{R})^2$$

3. We can stack the equations like we did for the Tikhonov regularization

$$\min_x \left\| \begin{matrix} b - Ax \\ \delta x - \delta x_0 \end{matrix} \right\|_2^2 = \min_x \left\| \begin{pmatrix} b \\ -\delta x_0 \end{pmatrix} - \begin{pmatrix} A \\ -\delta \end{pmatrix} x \right\|_2^2 =$$

$$= \|d - Cx\|_2^2 \Rightarrow C^TC\hat{x} = C^Td$$

$C = \begin{pmatrix} A \\ -\delta \end{pmatrix} I \qquad d = \begin{pmatrix} b \\ -\delta x_p \end{pmatrix}$

$$\begin{pmatrix} A \\ -\delta I \end{pmatrix}^T \begin{pmatrix} A \\ -\delta I \end{pmatrix} \hat{x} = \begin{pmatrix} A \\ -\delta I \end{pmatrix}^T \begin{pmatrix} b \\ -\delta x_p \end{pmatrix}$$

$$\Rightarrow (A^T - \delta I) \begin{pmatrix} A \\ -\delta I \end{pmatrix} \hat{x} = (A^T - \delta I) \begin{pmatrix} b \\ -\delta x_p \end{pmatrix}$$

$$\Rightarrow (A^TA + \delta^2 I)\hat{x} = A^Tb + \delta^2 I x_p$$

4. In class we saw that with random SVD, if A is low rank r then $QQ^TA \approx A$. We also saw that σ of $A \approx \sigma$ of Q^TA.

This gives us a much smaller rank 4 matrix that is now only 10×1000

5 $$E[\|X - X_0\|_F^2] = E[tr((X - X_0)^T(X - X_0))] =$$

$$= E[tr(x^Tx - x^Tx_0 - x_0^Tx + x_0^Tx_0)] = E[tr(x^Tx) - tr(x^Tx_0) - tr(x_0^Tx) + tr(x_0^Tx_0)]$$

because of trace linearity

$$= E[tr(x^Tx)] - 2tr(x_0^TE[X]) + tr(x_0^Tx_0) =$$

$$= E[\|x\|_F^2] - 2tr(x_0^Tx_0) + tr(x_0^Tx_0) = E[\|x\|_F^2] - \|x_0\|_F^2$$

definition of Frobenius Norm

6

a) We know that any positive-definite matrix W can be factored as $W = LL^T = U^TU$

$\|b - Ax\|_W = \sqrt{(b-Ax)^TW(b-Ax)} = \sqrt{(b-Ax)^TLL^T(b-Ax)} = \sqrt{(b^TLL^T - x^TA^TLL^T)(b-Ax)} =$

$= \sqrt{b^TLL^Tb - x^TA^TLL^Tb - b^TLL^TAx + x^TA^TLL^TAx}$

$\|b - Ax\|_W = \sqrt{(b-Ax)^TW(b-Ax)} = \sqrt{(b-Ax)^TU^TU(b-Ax)} = \sqrt{(U(b-Ax))^TU(b-Ax)}$

$= \|U(b-Ax)\|_2 = \|Ub - UAx\|_2 \Rightarrow \begin{matrix} d = Ub \\ C = UA \end{matrix} \Rightarrow \|d - Cx\|_2$

b) $C^TC\hat{x} = C^Td \Rightarrow (UA)^T(UA)\hat{x} = (UA)^TUb \Rightarrow A^TU^TUA\hat{x} = A^TU^TUb$

$\Rightarrow A^TWA\hat{x} = A^TWb$

7. We can re-write $b(a) = x_1 + ax_2$ as $b = Ax$ where $\underset{m \times 2}{A = \begin{bmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \vdots \\ 1 & a_m \end{bmatrix}}$, $\underset{2 \times 1}{x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}$, and $\underset{m \times 1}{b = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}}$. Since we know that b has independent random errors with equal variance σ^2 so we know $V^{-1} = I\frac{1}{\sigma^2}$. When we plug it into the more general equation, we get:

$$W = (A^TV^{-1}A)^{-1} = (A^T\tfrac{1}{\sigma^2}A)^{-1} = \sigma^2(A^TA)^{-1} = \sigma^2\begin{pmatrix} m & \sum^m a \\ \sum^m a & \sum^m a^2 \end{pmatrix}^{-1}$$

$$\Rightarrow \sigma^2 \frac{1}{det(A)}\begin{pmatrix} \sum^m a^2 & -\sum^m a \\ -\sum^m a & m \end{pmatrix} = \frac{\sigma^2}{m(\sum^m a^2 - (\sum^m a)^2)}\begin{pmatrix} \sum^m a^2 & -\sum^m a \\ -\sum^m a & m \end{pmatrix}$$

Variances of x are along the diagonal so

$$\text{variance}(x_1) = \frac{\sigma^2\sum a^2}{m(\sum^m a^2 - (\sum a)^2)} = \frac{\sigma^2\sum^m a^2}{\sum^m(a - \text{mean}(a))^2}$$

$$\text{Variance}(x_2) = \frac{\sigma^2 \cancel{m}}{\cancel{m}(\sum^m a^2 - (\sum^m a)^2)} = \frac{\sigma^2}{\sum^m(a - \text{mean}(a))^2}$$