LECTURE 9

Q. Given an overdetermined system $Ax=b$, what are the normal equations, and why do they give the solution to the least squares minimization problem?

§3.3 Perturbation Theory and Condition Number

defn. A = nonsingular matrix ($\Rightarrow \exists A^{-1}$)

The condition number of A (with respect to the operator norm $\|\cdot\|$) of A is $\kappa(A) = \|A\| \|A^{-1}\|$

eg. $\kappa_2(A) = \|A\|_2 \|A^{-1}\|_2$

& is a measure of how much the (unique) solution x_* to $Ax=b$ changes if both A and b are perturbed (ie add small change).

ie how sensitive the linear system is to a small perturbation; would the least squares solution change drastically if we had just a bit of noise in our data collection?

- $\kappa(A) = \kappa(A^{-1})$ (why?)
- large condition # means A is close to being singular

  "ill-conditioned" - ie not robust; may need diff model

  — actual "small" or "large" depends on your application! from following example, you'll get some intuition e.g. $A = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$ $\det(A) = 6 \Rightarrow A^{-1} = \frac{1}{6}\begin{pmatrix} 2 & -0 \\ -0 & 3 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$

$\|A\|_2 = \left[\max\left\{|\lambda_i(A^TA)|\right\}\right]^{1/2} = \sqrt{9} = 3$ $\|A^{-1}\|_2 = \left[\max\left\{\lambda_i((A^{-1})^T(A^{-1}))\right\}\right]^{1/2} = \sqrt{\frac{1}{4}} = \frac{1}{2}$

$A^TA = \begin{pmatrix} 4 & 0 \\ 0 & 9 \end{pmatrix}$ $(A^{-1})^T(A^{-1}) = \begin{pmatrix} \frac{1}{9} & 0 \\ 0 & \frac{1}{4} \end{pmatrix}$

$\Rightarrow K_2(A) = 3 \cdot \frac{1}{2} = \frac{3}{2}.$

② K_2 is the ratio b/w "λ_{max}" of A^TA and "λ_{min}" of A^TA.

(why?) $(A^{-1})^T = (A^T)^{-1}$.

$\Rightarrow \left((A^{-1})^T(A^{-1})\right)^T = (A^{-1})^T \cdot (A^{-1})^{TT}$

$= (A^T)^{-1}(A^{TT})^{-1} = (A^T \cdot A^{TT})^{-1} = (A^TA)^{-1}$

eigs(M) = eigs(M^T) where $M \in \mathbb{R}^{n\times n}$

(These $\lambda(A^TA)$ are the singular values we'll study later in the class.)

δA, δb represent a small change in A, b respectively.

We'll compare $Ax=b$ with $(A+\delta A)y = (b+\delta b)$

Thm 3.5 Assume A nonsingular and (choose δ such that)

$$\underbrace{\|\delta A\|}_{\text{a baby version of } A} \|A^{-1}\| = r < 1$$

Then the matrix $A+\delta A$ is still nonsingular, and

$$\|(A+\delta A)^{-1}\| \leq \frac{\|A^{-1}\|}{1-r}$$

The solution y of the perturbed system $(A+\delta A)y = (b+\delta b)$

satisfies $$\frac{\|y-x\|}{\|x\|} \leq \frac{\kappa(A)}{1-r}\left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|b\|}\right)$$

How to use:

- We won't go over proof. But the LHS is the rel error. This is bounded by the RHS, a function of κ, r, and δ.

  $\Rightarrow$ you choose threshold for your application.

Problems with normal equations method of finding opt sol to overdet system:

① Forming A^TA may lead to loss of information

② The condition number of A^TA is the square of that of A.

* pay attention – what are we doing with A^TA, A in diff contexts today?

Study Examples 3.12 (①) and 3.13 (②) in the book.