Lecture 11 — Numerical Analysis — 02/27/18

Last time:

Matrix conditioning: (For $A\underline{x} = \underline{b}$)

- $\|\underline{x} - \underline{x}'\| = \|A^{-1}\underline{b} - A^{-1}\underline{b}'\|$ — any norm

$$\leq \underbrace{\|A^{-1}\|}_{\text{Absolute condition number.}} \|\underline{b} - \underline{b}'\|$$

- $$\frac{\|\underline{x} - \underline{x}'\|}{\|\underline{x}\|} \leq \underbrace{\|A\| \, \|A^{-1}\|}_{\substack{\text{relative condition number.} \\ \underline{K(A)}}} \frac{\|\underline{b} - \underline{b}'\|}{\|\underline{b}\|}$$

In the 2-norm, recall that

$$\|A\|_2 = \max \sqrt{\lambda(A^T A)}$$ ← largest eigenvalue of $A^T A$.

⇒ If $A = U S V^T$ by SVD, then

$A^T A = V S^2 V^T$, ⇒ $\|A\|_2 = \sigma_1$ ← largest sing. value of A.

⇒ $A^{-1} = V S^{-1} U^T$ ⇒ $\|A^{-1}\|_2 = \frac{1}{\sigma_n}$ ← σ_n is smallest sing. val. of A.

$$\Rightarrow K_2(A) = \frac{\sigma_1}{\sigma_n}$$

Consequences of K_2: ~~Ex~~ Recall:

$$\frac{\|\underline{x}-\underline{x}'\|}{\|\underline{x}\|} \leq K_2 \frac{\|\underline{b}-\underline{b}'\|}{\|\underline{b}\|}$$

Imagine $\underline{b}'$ is the rounded version of $\underline{b}$, so then

$$\frac{\|\underline{b}-\underline{b}'\|}{\|\underline{b}\|} \sim O(\epsilon)$$

machine precision

$$\lesssim K_2 \epsilon .$$

$\underline{x}'$ is the solution to $A\underline{x}' = \underline{b}'$

This means that the "number of correct digits in $\underline{x}'$" is bounded by $-\log K_2 \cdot \epsilon$.

Ex: $\epsilon \sim 10^{-16}$

$K_2 \sim 10^{10}$

$$\Rightarrow \frac{\|\underline{x}-\underline{x}'\|}{\|\underline{x}\|} \leq 10^{-6}$$

$\Rightarrow$ we may ~~loose~~ lose $\log K_2$ digits in solving $A\underline{x} = \underline{b}$, this does not mean that ~~the~~ ~~be~~ $\underline{r} = \underline{b} - A\underline{x}'$ is large.

Depends on Backward stability

The SVD

The factorization $A = USV^T$ exists for any matrix A, with U, V orthogonal and $S = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$ (obviously real.).

Proof by construction:

Let $\underline{v}_j$ be the eigenvectors of A^TA. (assume n distinct $\lambda_j > 0$ for now.)

Then $$(A\underline{v}_i, A\underline{v}_j) = (\underline{v}_i, A^TA\underline{v}_j) = (\underline{v}_i, \lambda_j \underline{v}_j) = \lambda_j \underbrace{(\underline{v}_i, \underline{v}_j)}$$

$= 0$ since A^TA is real-symmetric (HW question).

This means that $(A\underline{v}_i, A\underline{v}_j) = 0$, i.e. $A\underline{v}_i$, $A\underline{v}_j$ are orthogonal, Each is in the column space of A (obviously), and therefore form a basis for $\mathrm{col}(A)$.

$$\mathrm{col}(A) = \mathrm{span}\{A\underline{v}_1, \ldots, A\underline{v}_n\}.$$

Also, $\|A\underline{v}_i\| = (A\underline{v}_i, A\underline{v}_i) = (\underline{v}_i, \lambda_i \underline{v}_i) = \lambda_i$.

Set $\underline{u}_j = \frac{1}{\sqrt{\lambda_j}} A\underline{v}_j$.

Set the matrix:

$$U = \begin{pmatrix} \underline{u}_1 & \underline{u}_2 & \cdots & \underline{u}_n \end{pmatrix}$$

$$V = \begin{pmatrix} \underline{v}_1 & \underline{v}_2 & \cdots & \underline{v}_n \end{pmatrix}$$

$$S = \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix} \qquad \text{(sorted)}$$

Then trivially,

$$\begin{aligned} AV &= A\begin{pmatrix} \underline{v}_1 & \underline{v}_2 & \cdots & \underline{v}_n \end{pmatrix} \\ &= \begin{pmatrix} A\underline{v}_1 & \cdots & A\underline{v}_n \end{pmatrix} = US \\ &= \begin{pmatrix} \frac{A\underline{v}_1}{\sqrt{\lambda_1}} & \cdots & \frac{A\underline{v}_n}{\sqrt{\lambda_n}} \end{pmatrix} \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix} \\ &= \begin{pmatrix} A\underline{v}_1 & \cdots & A\underline{v}_n \end{pmatrix} \end{aligned}$$

so $A = USV^T$.

What if A is not full rank? $\dim \operatorname{col}(A) < n$.

<u>Truncated SVD</u>

Then $A = U_{n\times k} \left(\Sigma_{k\times k} \right) V^T_{k\times n}$

We will discuss methods for computing the SVD after eigenvalues, least squares. (computing A^TA is not recommended!)

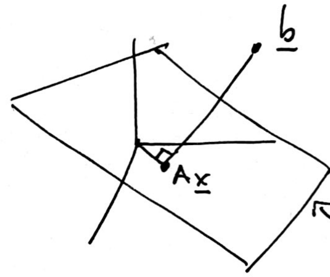
Next: Least squares:

Often, A is not square, but we would still like to find the "best" solution to $A\underline{x} = \underline{b}$.

Least squares: For $A \in \mathbb{R}^{m \times n}$, $m > n$, find $\underline{x}$ such that $\|A\underline{x} - \underline{b}\|_2$ is as small as possible.

important

Geometrically:



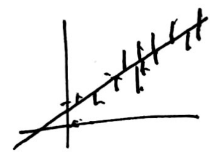
minimize distance between $\underline{b}$ and $A\underline{x}$.

We have that $\|A\underline{x} - \underline{b}\|^2$ (2-norm notation supressed):

$$= \sum_{i=1}^{m} \left(b_i - \sum_{j=1}^{n} a_{ij} x_j \right)^2$$

Many other interpretations/applications:

- calculus
- statistics



You probably learned before that least squares problems can be solved by instead solving:

$$\underline{A^T A} \underline{x} = A^T \underline{b} \quad \text{(the normal equation)}$$

but this <u>squares</u> the condition number!

(Aside: the normal equations can be obtained by computing the minimum of $\| A\underline{x} - \underline{b} \|^2$ by taking partial derivatives, etc.

Also: $\underline{r} = \underline{b} - A\underline{x}$

If $\underline{x}$ is the least square solution, then $\underline{r} \perp \text{col} A$, in particular $A^T \underline{r} = \underline{0}$.

$$\Rightarrow A^T \underline{r} = 0 = A^T \underline{b} - A^T A \underline{x}$$

$$\Rightarrow A^T A \underline{x} = A^T \underline{b} \quad \text{has at } \underline{\text{least}} \text{ one solution.}$$)

How else can least squares be solved? Another matrix factorization... QR.