Last time: Linear equation $Ax = b$

$A \in \mathbb{C}^{n\times n}$, $b \in \mathbb{C}^n$, A invertible

Condition # of A $\quad \kappa(A) = \|A\| \, \|A^{-1}\|$

When A is a normal matrix w/ eig. val.

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n| > 0$$

then $\kappa_2(A) = \|A\|_2 \, \|A^{-1}\|_2 = \dfrac{|\lambda_1|}{|\lambda_n|}$

$\kappa_2(A)$ can be large when $|\lambda_1| \gg |\lambda_n|$ (ill-conditioned problem)

- For ill-conditioned problems, forward error can be large even when the algorithm is backward stable

  How do we decrease the condition number?

  (regularize the problem)

- For simplicity, consider A being real symmetric positive definite w/ eig. val. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$

we may solve the following regularized problem

$$(A + \alpha I)\hat{x} = b \qquad (\alpha > 0)$$

Tikhonov regularization

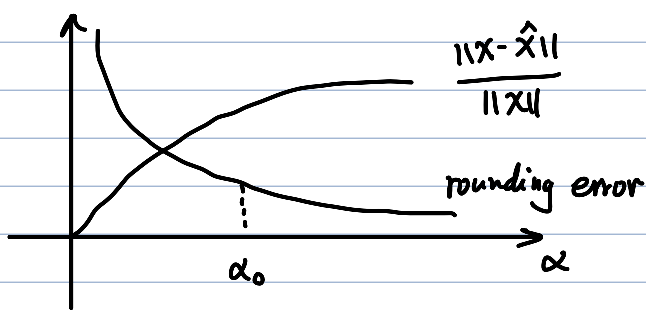
$$\kappa_2(A) = \frac{\lambda_1 + \alpha}{\lambda_n + \alpha} < \kappa_2(A)$$

(If A is not s.p.d. or even just a full column rank rectangular matrix, we can solve

$$(A^*A + \alpha I)x = A^*b$$

which is equivalent to $\min_x \|b - Ax\|^2 + \alpha\|x\|^2$)

- How do we choose α?



Since

$$x-\hat{x} = (A+\alpha I)^{-1}(A+\alpha I - A)A^{-1}b$$
$$= \alpha(A+\alpha I)^{-1}x$$

we have $\|x-\hat{x}\|_2 \le \alpha\|(A+\alpha I)^{-1}\|_2\|x\|_2 = \frac{\alpha}{\lambda_n+\alpha}\|x\|_2$

we hope to balance $\frac{\|x-\hat{x}\|_2}{\|x\|_2} \sim \kappa_2(A+\alpha I)\,\varepsilon_{mach}$

that is $\frac{\alpha}{\lambda_n+\alpha} \sim \frac{\lambda_1+\alpha}{\lambda_n+\alpha}\varepsilon_{mach}$

that is $\alpha \sim \lambda_1 \varepsilon_{mach}$

- Tikhonov regularization is useful in solving inverse problems where the problem is usually ill-conditioned (a small perturbation in input can lead to a non-negligible change in output)
- An classical textbook on this topic is "An intro. to the mathmatical theory of inverse problems" by A. Kirsch.

Last time: SVD of matrix $A \in \mathbb{C}^{m\times n}$

$$A = U\ \Sigma\ V^*$$

$U \in \mathbb{C}^{m\times m}$, $U^*U = I_m$

$\Sigma \in \mathbb{R}^{m\times n}_{\ge 0}$, $\Sigma = \mathrm{diag}(\sigma_1,\cdots,\sigma_p)$, $p=\min\{m,n\}$

$V \in \mathbb{C}^{n\times n}$, $V^*V = I_n$

- Existence: ① $A^*A = V\underbrace{\begin{bmatrix}\sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2\end{bmatrix}}_{\Sigma^2}V^*$

② $U\Sigma = AV$

- Practical algorithm: Solve eigenvalue problem for

$$H = \begin{bmatrix} 0 & A^* \\ A & 0 \end{bmatrix} \leftarrow \text{Hermitian}$$

use Golub-Kahan bidiagonalization: operation count $\approx 4mn^2$ FLOPs

---

- Application : Overdetermined least-squares

Given $b(a) \in L^2([0,1])$

Find n-degree polynomial $p(a) \in P_n([0,1])$

such that $$\min_{p \in P_n} \int_0^1 |p(a) - b(a)|^2 \, da$$

Discretize the problem:

quadrature: $\int_0^1 f(a)\, da \approx \frac{1}{m} \sum_{i=1}^{m} f(a_i)$, $\quad a_i = \frac{i-1}{m-1}, \quad i = 1, \dots, m$

$$p(x) = \sum_{k=0}^{n-1} x_k a^k$$

$$\Rightarrow \quad \min_{c \in \mathbb{R}^n} \frac{1}{m} \sum_{i=1}^{m} \left| \sum_{k=0}^{n-1} x_k a_i^k - b(x_i) \right|^2$$

Let $$A = \begin{bmatrix} 1 & a_1 & \cdots & a_1^{n-1} \\ 1 & a_2 & \cdots & a_2^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & a_m & \cdots & a_m^{n-1} \end{bmatrix}, \quad b = (b(x_1), \dots, b(x_m))^T$$

$$x = (x_1, \dots, x_n)^T$$

$$\Rightarrow \quad \min_{x} \|Ax - b\|_2^2 \quad \leftarrow \text{Least-squares problem}$$

m equations $>$ n variables

$$\begin{array}{cc} & n \\ m & \begin{bmatrix} \quad & \quad \\ & \\ & \\ & \end{bmatrix} \end{array} \begin{bmatrix} \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix}$$

A $\quad x \quad$ b

How to solve least squares?

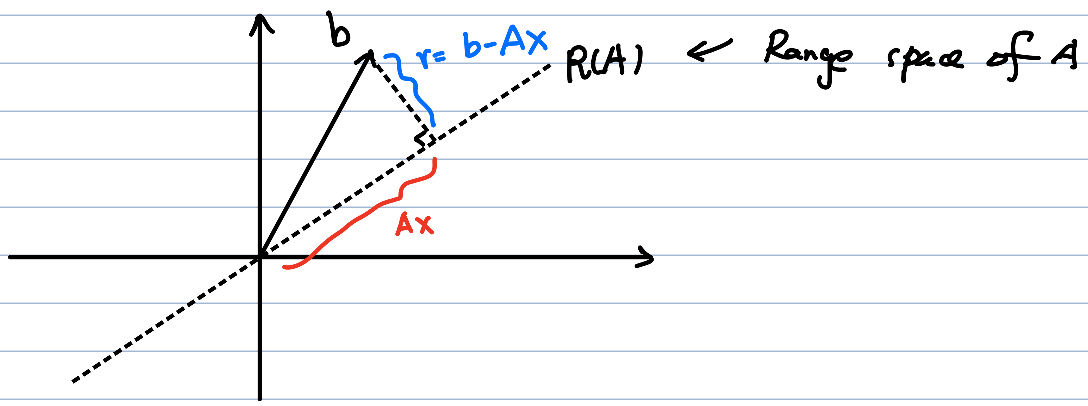
- via normal equations

$$\hat{x} \in \arg\min_x \|Ax-b\|_2^2$$

$$\iff \quad r = b - A\hat{x} \perp R(A)$$

$$\iff \quad b - A\hat{x} \in N(A^*)$$

$$\iff \quad A^*A\hat{x} = A^*b \quad \longleftarrow \text{ Normal equations}$$



assume $\operatorname{rank}(A) = n \implies \operatorname{rank}(A^*A) = \operatorname{rank}(A) = n \implies A^*A$ invertible

- Solving least-squares via normal equations

Step 1: compute $C = A^*A$, $d = A^*b$

Step 2: Solve $C\hat{x} = d$ $\qquad$ operation count $\approx 2mn^2$ FLOPs $(m \gg 1)$

Remark: Solving normal equations directly can suffer from ill-conditioning

For the least square problem using monomials,

as $m \to +\infty$, $\frac{1}{m}(A^*A)_{ij} = \frac{1}{m}\sum_{k=1}^{m} a_k^{i+j-2}$

$$= \frac{1}{m}\sum_{k=1}^{m}\left(\frac{k-1}{m-1}\right)^{i+j-2}$$

$$\to \int_0^1 a^{i+j-2}\,da = \frac{1}{i+j-1}$$

So $\frac{1}{m}A^*A \to H_n$ Hilbert matrix – ill-conditioned!

In general, $\kappa_2(A^*A) = \frac{\sigma_1^2}{\sigma_n^2} >> \sigma_1/\sigma_n \; (= \kappa_2(A))$

Forward error $\approx$ $\kappa_2^2(A)\,\varepsilon \leftarrow$ backward error

Remark: monomials are "bad" basis, choose basis that are orthonormal (Legendre polynomial)

- Solve Least-squares via SVD

Let $A = U\Sigma V^*$

$$\min_x \| U\Sigma V^* x - b\|_2^2$$

Let $d = U^*b$, $y = V^*x$

$$\min_y \|\Sigma y - d\|_2^2 = \min_{y\in\mathbb{C}^n} \left\| \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ & 0 & \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} - \begin{bmatrix} d_1 \\ \vdots \\ d_m \end{bmatrix} \right\|^2$$

Let $\Sigma^+ = \begin{bmatrix} d_1/\sigma_1 & & & \\ & \ddots & & \\ & & d_n/\sigma_n & 0 \end{bmatrix}$

$$\hat{y} := \Sigma^+ d = \arg\min \|\Sigma y - d\|_2^2$$

hence $\hat{x} = V\hat{y} = \underbrace{V\Sigma^+U^*}_{A^+} b \in \arg\min \|Ax-b\|^2$

A^+ pseudoinverse of A

Actually, from normal equation,

$$\hat{x} = \underbrace{(A^*A)^{-1}A^*}_{=A^+} b$$

$$= (V\Sigma^*\Sigma V^*)^{-1} V\Sigma^*U^*b$$

$$= V(\Sigma^*\Sigma)^{-1}V^*V\Sigma^*U^*b$$

$$= V\Sigma^+ U^* b$$

Algorithm:

Step 1: Compute SVD: $A = U\Sigma V^*$

operation count $\approx 4mn^2$ FLOPs

Step 2: $\hat{x} = V\Sigma^+ U^* b$

- Stability of least-squares

Define $\kappa_2(A) = \|A\|_2 \|A^+\|_2 = \sqrt{\kappa_2(A^*A)}$

<u>Thm (Wedin, 1973)</u>

Let $A \in \mathbb{C}^{m\times n} (m \geq n)$. Assume $A, A+\Delta A$ both be of full column rank.

Let $x = \underset{y\in\mathbb{C}^n}{\operatorname{argmin}} \|b - Ay\|_2$. $r = b - Ax$

and $\hat{x} = \underset{y\in\mathbb{C}^n}{\operatorname{argmin}} \|(b+\Delta b) - (A+\Delta A)y\|_2$, $s = b+\Delta b - (A+\Delta A)\hat{x}$

where $\|\Delta A\|_2 / \|A\|_2$, $\|\Delta b\|_2 / \|b\|_2 \leq \varepsilon$

Then provided that $\kappa_2(A)\varepsilon < 1$

we have

$$\frac{\|\hat{x}-x\|_2}{\|x\|_2} \leq \frac{\kappa_2(A)\varepsilon}{1-\kappa_2(A)\varepsilon}\left(2 + (\kappa_2(A)+1)\frac{\|r\|_2}{\|A\|_2\|x\|_2}\right)$$

$\begin{cases} \approx \kappa_2(A)\varepsilon & \text{when } \|r\|_2 \text{ small} \\ \approx \kappa_2^2(A)\varepsilon & \text{otherwise} \end{cases}$

$$\frac{\|r-s\|_2}{\|b\|_2} \leq (1+2\kappa_2(A))\,\varepsilon_{\text{mach}}$$

These bounds are approximately attainable

Pf: See Higham Thm 20.1 ☒

Remark: The first bound is usually interpreted as saying that the sensitivity of least-squares is measured by $\kappa_2(A)$

when $\|r\|$ is small or zero and by $K_2^2(A)$ otherwise

<u>Thm</u> Solving least-squares via SVD is backward stable.
The computed $\hat{x}$ satisfies

$$\hat{x} = \underset{x}{\arg\min} \|(A+\Delta A)x + b + \Delta b\|_2$$

and $\|\Delta A\| / \|A\|$, $\dfrac{\|\Delta b\|}{\|b\|} \leq C_{m,n}\, \varepsilon_{mach}$

Summary:

method	normal eqn	SVD
conditioning	$\approx K_2^2(A)$	$\approx K_2(A)$ ($\|r\|$ small)
operation count	$\approx 2mn^2$ FLOPs	$\approx 4mn^2$ FLOPs

- Solving least-squares via QR factorization

  Computing SVD is expensive. can we work with orthogonal transform but lower cost?

- (reduced) QR factorization

$$A = Q\;R$$

$R \in \mathbb{C}^{n\times n}$
upper triangular matrix

$Q \in \mathbb{C}^{m\times n}$
$Q^*Q = I_n$
orthonormal columns $\qquad (m \geq n)$

$\boxed{\text{Not } QQ^* = I_m}$

when $\operatorname{rank}(A) = n$, $\Rightarrow \operatorname{rank}(R) = n$

hence $R(A) = R(Q)$

Let $Q = [\, q_1 \cdots q_n \,]$, $q_i \in \mathbb{C}^m$

projection onto $R(A)$

Then $\quad P_A x = P_Q x = \sum_{i=1}^{n} q_i (q_i^* x) = Q Q^* x$

Hence $\quad \hat{x} \in \arg\min_x \|Ax - b\|_2^2$

$$\iff \quad r = b - A\hat{x} \perp R(A)$$

$$\iff \quad P_Q r = 0$$

$$\iff \quad QR\hat{x} = QQ^* b$$

$$\iff \quad R\hat{x} = Q^* b$$

$$\iff \quad \hat{x} = R^{-1} Q^* b$$

Actually, from normal equations,

$$\hat{x} = (A^*A)^{-1} A^* b = (R^*R)^{-1} R^* Q^* b = R^{-1} Q^* b$$