Last time: $f(x): \mathbb{R}^n \to \mathbb{R}^m$

The condition number $\kappa_f(x) := \dfrac{\|Df(x)\| \, \|x\|}{\|f(x)\|}$

Today: Apply it to $Ax = b$

- Motivation: $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$, A invertible

Solving $Ax=b$ exactly: $f(b) = A^{-1}b$

A backward stable algorithm satisfies $\hat{f}(b) = f(\tilde{b}) = A^{-1}(b+\Delta b)$
with $\|\Delta b\| / \|b\| = O(\varepsilon_{mach})$

To analyze forward error, need to compute κ_A

$$\kappa_A(b) = \frac{\|Df\| \, \|b\|}{\|A^{-1}b\|} = \frac{\|A^{-1}\| \, \|b\|}{\|A^{-1}b\|}$$

We may want a condition number that is independent of input b,

$$\kappa(A) := \sup_{b\in\mathbb{R}^n} \frac{\|A^{-1}\| \, \|b\|}{\|A^{-1}b\|} = \|A^{-1}\| \, \|A\|$$

$\|b\| = \|AA^{-1}b\| \le \|A\| \, \|A^{-1}b\|$

"=" attainable for some $b \in \mathbb{R}^n$

$\kappa(A)$ is called the condition number of A

- Properties of condition #s (Assume $\|\cdot\|$ is subordinate norm in the following)

1) $\kappa(A) \geq 1$, $\kappa(A^{-1}) = \kappa(A)$

2) If A is normal with eigenvalues $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n| > 0$

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

For general A, we have $\kappa(A) \geq \dfrac{|\lambda_1|}{|\lambda_n|}$

Pf: When A is normal, $A = U^* \Lambda U$.

$$\|A\|_2 = \sqrt{\lambda_{max}(A^* A)} = \sqrt{\lambda_{max}(U^* \Lambda^2 U)} = |\lambda_1|$$

Similarly $\|A^{-1}\|_2 = |\lambda_n|^{-1}$

For general A, we need the following relation between norm and maximun eigenvalue

Lemma: The spectral radius of A, denote by

$$\rho(A) = \{ |\lambda| : \lambda \text{ is an eigenvalue of } A \}$$

satisfies $\rho(A) \leq \|A\|$

Proof of lemma: exercise $\blacksquare$

Using this lemma, we have $\|A\| \geq |\lambda_1|$, $\|A^{-1}\| \geq |\lambda_n|^{-1}$ and the lower bound of A follows $\blacksquare$

3) condition #s of different norms are equivalent.

In particular, $\frac{1}{n} K_2(A) \leq K_1(A) \leq n K_2(A)$

4) 2- condition # measures the inverse of the relative distance to the nearest singular matrix.

Prop. Let $A \in \mathbb{R}^{n \times n}$, A not singular, then

$$\frac{1}{K_2(A)} = \min_{\substack{\delta A:\, A+\delta A \\ \text{singular}}} \left\{ \frac{\|\delta A\|_2}{\|A\|_2} \right\}$$

Pf: exercise. $\blacksquare$

We have the following thm that tells you how the solution is perturbed given an inaccurate right hand side.

Thm 1 Let $Ax = b$ and $r = b - A\hat{x}$ ← residual

then we have

$$\frac{1}{K(A)} \frac{\|r\|}{\|b\|} \leq \frac{\|\hat{x} - x\|}{\|x\|} \leq K(A) \frac{\|r\|}{\|b\|}$$

Pf: Let $\Delta x = \hat{x} - x$

$$\|\Delta x\| = \|A^{-1} r\| \leq \|A^{-1}\| \, \|r\|$$

use $\|b\| = \|Ax\| \leq \|A\| \, \|x\|$ we obtain inequ on the right.

The left inequ can be proved similarly ← exercise ☒

$K(A)$ seems to only quantify the error in output for a perturbation in b, but it turns out that it also quantify the perturbation in A

Thm Let $Ax = b$, $(A + \Delta A)\hat{x} = b + \Delta b$

form of eqn we get in backward error analysis of many algo for solving $Ax=b$

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

and $\varepsilon K(A) < 1$. then

$$\frac{\|x - \hat{x}\|}{\|x\|} \leq \frac{K(A)}{1 - K(A) \cdot \|\Delta A\| / \|A\|} \left(\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta b\|}{\|b\|} \right)$$

$$= O\left(K(A)\, \varepsilon\right) \quad \text{for} \quad \varepsilon \ll \frac{1}{K(A)}$$

Pf: $\hat{x} = (A + \Delta A)^{-1}(b + \Delta b)$

$x = (A + \Delta A)^{-1}(A + \Delta A)\, x = (A + \Delta A)^{-1}(b + (\Delta A) x)$

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

$$= [A(I + A^{-1}\Delta A)]^{-1}(\Delta b - (\Delta A)x)$$

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

$$\|\hat{x} - x\| \leq \|(I + A^{-1}\Delta A)^{-1}\| \, \|A^{-1}\| \, (\|\Delta b\| + \|\Delta A\| \cdot \|x\|)$$

we use $\|b\| = \|Ax\| \leq \|A\| \, \|x\|$

$$\frac{\|\hat{x} - x\|}{\|x\|} \leq \|(I + A^{-1}\Delta A)^{-1}\| \, \kappa(A) \left(\frac{\|\Delta b\|}{\|b\|} + \frac{\|\Delta A\|}{\|A\|}\right) \quad (*)$$

Lemma: If $\|B\| < 1$, then $I + B$ is invertible

and $\|(I+B)^{-1}\| \leq \dfrac{1}{1 - \|B\|}$

Pf: exercise $\boxed{}$

Using this lemma, we have

$$\|(I + A^{-1}\Delta A)^{-1}\| \leq \frac{1}{1 - \|A^{-1}\Delta A\|} \leq \frac{1}{1 - \kappa(A) \|\Delta A\| / \|A\|} \quad (**)$$

The proof is finished after combining $(*)$ $(**)$ $\boxed{}$

---

ex. Matrix with large condition number (ill-posed problem)

Hibert matrix $H_n = \left(\dfrac{1}{i+j-1}\right)_{1 \leq i,j \leq n}$

$$\kappa_2(H_n) = \|H_n\|_2 \, \|H_n^{-1}\|_2$$

$$\approx \frac{(1+\sqrt{2})^{4n}}{\sqrt{n}} \approx \frac{34^n}{\sqrt{n}}$$

when $n = 20$, $\kappa_2(H_n) \approx 10^{29} >> 10^{16} = 1/\varepsilon_{mach}$

Review : unitary matrix $Q^*Q = I$, $Q \in \mathbb{C}^{n\times n}$

$Q = [\ q_1 \cdots q_n]$, $q_i \in \mathbb{C}^n$, $i = 1, \cdots, n$

Q is unitary $\iff$ $q_i^* q_j = \delta_{ij}$, $i,j = 1, \cdots, n$

i.e. $\{q_1, \cdots, q_n\}$ forms an orthonormal basis on $\mathbb{C}^n$

Some properties: Let $x \in \mathbb{R}^n$, $A \in \mathbb{C}^{n\times n}$

$$\left.\begin{aligned} \|Qx\|_2 &= \|x\|_2 \\ \|QA\|_2 = \|AQ\|_2 &= \|A\|_2 \\ \|QA\|_F = \|AQ\|_F &= \|A\|_F \end{aligned}\right\} \text{unitary invariant norm}$$

Pf: $\|Qx\|_2^2 = x^*Q^*Qx = x^*x = \|x\|_2^2$

$\|QA\|_2^2 = \sqrt{\lambda_{max}(A^*Q^*QA)} = \sqrt{\lambda_{max}(A^*A)} = \|A\|_2^2$

$\|QA\|_F^2 = \sqrt{tr(A^*Q^*QA)} = \sqrt{tr(A^*A)} = \|A\|_F^2$

Other equation can be proved similarly $\leftarrow$ exercise ▨

In Linear algebra course, we know when $A \in \mathbb{C}^{n\times n}$ is normal, then A is diagonalizable, i.e., can be factored as

$$A = U \Lambda U^*$$

where $U \in \mathbb{C}^{n\times n}$ is unitary, $\Lambda = diag(\lambda_1, \cdots, \lambda_n)$

- Q: A is not normal? $A \in \mathbb{C}^{m\times n}$?

- Singular value Decomposition

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

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

where $U = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{bmatrix} \in \mathbb{C}^{m\times m}$, u_i left singular vector

$V = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix} \in \mathbb{C}^{n\times n}$. v_j right singular vector

$\Bigg\}$ unitary

$\Sigma = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_p \\ & & \end{bmatrix} \in \mathbb{R}^{m\times n}$, $p = \min\{m,n\}$ diagonal

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$ singular values

it's not always possible to make $u_i^* v_j = \delta_{ij}$

Existence of SVD:

Since A^*A is Hermitian, positive-definite,

we have $A^*A = V \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{bmatrix} V^*$

i.e. $A^*A v_i = \sigma_i^2 v_i$, $i = 1, \cdots, n$, $v_i^* v_j = \delta_{ij}$

Since $\operatorname{rank}(A^*A) = \operatorname{rank}(A) =: r$

we know $\sigma_1 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_n = 0$

Let $u_i = \dfrac{A v_i}{\sigma_i}$, $i = 1, \cdots, r$. $u_i \in \mathbb{C}^m$

Clearly $u_i^* u_j = \dfrac{v_i^* A^* A v_j}{\sigma_i \sigma_j} = \delta_{ij}$, $i,j = 1, \cdots, r$

and $A v_i = \sigma_i u_i$, $i = 1, \cdots, r$ by definition.

Since $\dim(\operatorname{Null}(A)) = n - r$, we know $A v_i = 0$, $i = r+1, \cdots, m$

We can enlarge u_i to a set of orthonormal basis

$\{u_1, \cdots, u_r, u_{r+1}, \cdots, u_m\}$, s.t. $u_i^* u_j = \delta_{ij}$. $i,j = 1, \cdots, m$ ▨

Facts about SVD:

1) $r := \text{rank}(A) = \#$ non-zero SV's σ_i

2) $\text{range}(A) = \text{span}\{u_1, \cdots, u_r\}$

$\text{null}(A) = \text{span}\{v_{r+1}, \cdots, v_n\}$

3) $\|A\|_2 = \sigma_1$,

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

$$\kappa_2(A) = \sigma_1 / \sigma_n$$

Pf: $\|A\|_2 = \|U\Sigma V^*\|_2 = \|\Sigma\|_2 = \sqrt{\lambda_{\max}(\Sigma^*\Sigma)} = \sigma_1$

Similarly $\|A^{-1}\|_2 = \sigma_n^{-1}$ and hence $\kappa_2(A) = \sigma_1/\sigma_n$

4) $A = \sum_{i=1}^{r} \sigma_i u_i v_i^*$,

Best rank k approximation: $A_k := \sum_{i=1}^{k} \sigma_i u_i v_i^*$ $(k \ll m,n)$

$$\min_{\text{rank}(B)\le r} \|A - B\|_F = \|A - A_k\|_F = \sqrt{\sum_{i>k} \sigma_i^2} \quad (1)$$

$$\min_{\text{rank}(B)\le r} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1} \quad (2)$$

(Eckart-Young, 1936)

5) Point cloud alignment

Many tasks in computer vision involve the alignment of 3D shapes. Suppose we have a laser scanner that collects two point clouds of the same rigid objects from different views. how can we align these two point clouds into a single coordinate frame.

Consider two sets of point clouds

$$X = [\, x_1 \ \cdots \ x_n \,], \quad Y = [\, y_1 \ \cdots \ y_n \,] \in \mathbb{R}^{d \times n}$$

We aim to solve the following orthogonal Procrustes problem.

$$\min_{\substack{R:\, R^T R = I \\ R \in \mathbb{R}^{d \times d}}} \| RX - Y \|_F^2$$

To find R, we note that

$$\| RX - Y \|_F^2 = tr\left[(RX - Y)^T (RX - Y) \right]$$

$$= tr\left(X^T X - Y^T R X - X^T R^T Y + Y^T Y \right)$$

Thus, we want to maximize

$$tr(Y^T R X + X^T R^T Y) = 2\, tr(Y^T R X) = 2\, tr(R X Y^T)$$

Let $XY^T = U \Sigma V^T$, then

$$tr(R X Y^T) = tr(R U \Sigma V^T)$$

$$= tr(\underbrace{V^T R U}_{\tilde{R}} \Sigma)$$

$$= \sum_{i=1}^{d} \tilde{r}_{ii} \sigma_i$$

Since $\tilde{R}$ is orthogonal, $|\tilde{r}_{ii}| \le 1$, and hence

$$tr(R X Y^T) \le \sum_{i=1}^{d} \sigma_i$$

"=" iff $\tilde{r}_{ii} = 1$, again since $\tilde{R}$ is orthogonal, $\tilde{R} = I$

hence $R = V U^T$

b) Regularizing ill-posed problems

Consider solving the following problem

$$Ax = b \quad \text{with} \quad A \in \mathbb{R}^{n\times n}, \quad \lambda_i(A) > 0$$

$$K_2(A) = \frac{\lambda_{max}(A)}{\lambda_{min}(A)} \gg 1$$

Since the problem is ill-posed, any algorithm that solves the problem is difficult.

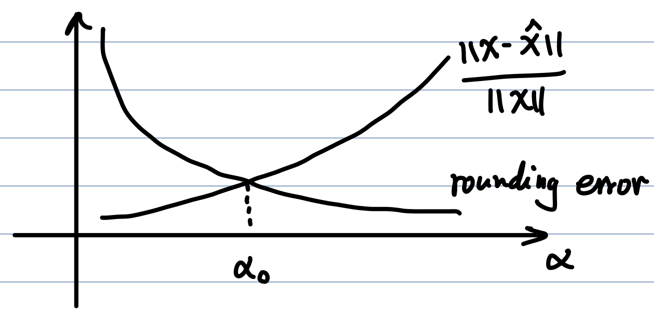
How do we decrease the condition number?
(regularize the problem)

we may solve instead

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

$$K_2(A) = \frac{\lambda_{max}(A) + \alpha}{\lambda_{min}(A) + \alpha} < K_2(A)$$

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 $\|x-\hat{x}\|_2 \leq \alpha \|(A+\alpha I)^{-1}\|_2 \|x\|_2 = \dfrac{\alpha}{\lambda_{min}(A) + \alpha} \|x\|_2$

we hope to balance $\dfrac{\|x-\hat{x}\|_2}{\|x\|_2} \sim K_2(A+\alpha I)\, \varepsilon_{mach}$

that is $$\frac{\alpha}{\lambda_{min}(A)+\alpha} \sim \frac{\lambda_{max}(A)+\alpha}{\lambda_{min}(A)+\alpha}\,\varepsilon_{mach}$$

that is $$\alpha \sim \lambda_{max}(A)\,\varepsilon_{mach}$$