Welcome to Ma221! (Mar 15)

Perturbation Theory

Can I trust my answer?

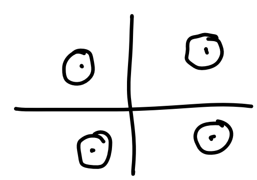
Last time: $A = I$ showed eigenvectors very sensitive: takeaway. close evals $\Rightarrow$ evecs ill conditioned

To describe perturbations in evals:

Def: Epsilon-pseudo-spectrum of A is set of all evals of all matrices within distance ε of A:

$$\Lambda_\varepsilon(A) = \{ \lambda : (A+E)x = \lambda x \text{ for some } x \neq 0,\ \|E\|_2 \le \varepsilon \}$$

smallest possible $\Lambda_\varepsilon(A)$: each disk around eval of A, radius ε



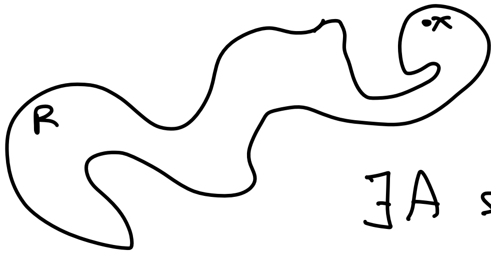
attained by $E = \varepsilon I$

attained by $A = A^H$ (Chap 5)

Worst (most sensitive) case

Thm (Trefethen + Reichel)

given any simply connected $R \subseteq \mathbb{C}$
(no holes)



any $x \in R$
any $\varepsilon > 0$

$\exists A$ s.t. $\Lambda_\varepsilon(A)$ fills out R

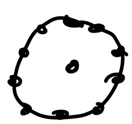
proof: Ma 185 (Riemann Mapping Thm)

Ex: Perturb $n \times n$ Jordan Block, $\lambda = 0$
with $J(n,1) = \varepsilon$

$$\begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ \varepsilon & & & 0 \end{bmatrix}$$

$p(\lambda) = \lambda^n - \varepsilon$

$\Rightarrow \lambda = \sqrt[n]{\varepsilon}$ uniformly spaced on circle of radius $|\sqrt[n]{\varepsilon}|$



$\varepsilon = 10^{-16}$ $\quad n = 16$

$\sqrt[n]{\varepsilon} = .1$

(1) evals are continuous but not necessarily differentiable
(slope of $\sqrt[n]{\varepsilon} = \infty$ at $\varepsilon = 0$

(2) expect sensitive evals when evals nearly multiple

Condition number of simple (nonmultiple) evals (else ∞)

Thm: λ be simple eval of A

$$Ax = \lambda x \quad , \quad y^H A = \lambda y^H \qquad \|x\|_2 = \|y\|_2 = 1$$

If we perturb A to $A+E$

λ perturbed to $\lambda + \delta\lambda$

$$\delta\lambda = \frac{y^H E x}{y^H x} + O(\|E\|^2)$$

$$|\delta\lambda| \le \frac{\|E\|_2}{|y^H x|} + O(\|E\|^2)$$

$$= \sec(\theta) \cdot \|E\|_2 + O(\|E\|^2)$$

θ = angle between x and y

$\sec(\theta)$ = condition number

proof: $(A+E)(x+\delta x) = (\lambda + \delta\lambda)(x + \delta x)$

$$\underbrace{Ax}_{\text{cancel}} + A\delta x + Ex + \underbrace{E\delta x}_{\text{second order drop}} = \underbrace{\lambda x}_{\text{cancel}} + \lambda\delta x + \delta\lambda\, x + \underbrace{\delta\lambda\,\delta x}_{\text{second order drop}}$$

$$y^H (A\delta x + Ex = \lambda \delta x + \delta\lambda \cdot x)$$

$$\underbrace{y^H A \delta x}_{\text{cancel}} + y^H E x = \underbrace{y^H \lambda \delta x}_{\text{cancel}} + y^H \delta\lambda\, x$$

$$\frac{y^H E x}{y^H x} = \delta\lambda$$

Special Case 1: $A = A^H$ or "normal"

$AA^H = A^H A$ (HW Q 4.2)

$\Rightarrow$ A has orthonormal evecs

Cor: If A normal, perturbing A to $A+E$

$\Rightarrow |\delta\lambda| \le \|E\|_2 + O(\|E\|^2)$

i.e. condition # = 1

proof: $A = Q \Lambda Q^T$

$AQ = Q\Lambda$ and $Q^T A = \Lambda Q^T$

right evecs = cols of Q

left evecs = rows of Q^H = cols of Q

$\Rightarrow x = y \Rightarrow y^H x = 1$

Later (Chap 5) if $A = A^H$ and $E = E^H$

then $|\delta\lambda| \le \|E\|_2$ (no $\|E\|^2$ term)

Special Case 2: A = Jordan Block

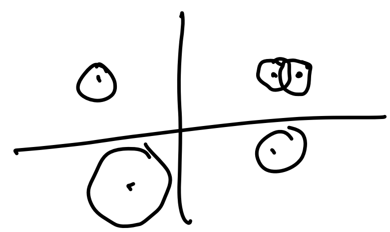
$$A = \begin{bmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix} \quad x = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad y = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

$y^H x = 0 \Rightarrow$ cond # = ∞

Extend to eliminate $O(\|E\|^2)$ term

Thm (Bauer-Fike) if A has all simple evals, λ_i, with right and left evecs x_i and y_i, $\|x_i\|_2 = \|y_i\|_2 = 1$

Then for any E, the evals of $A+E$ lie in disks in $\mathbb{C}$, centered at λ_i with radii $n \cdot \frac{\|E\|_2}{|y_i^H x_i|}$



if k disks overlap, k evals lie in their union (proof in book)

Algorithms for NonSymmetric Eigenproblem

Ultimate Algorithm:

Hessenberg QR (HQR)

Takes any $n \times n$ nonsymmetric dense A computes Schur form $A = Q^H T Q$ in $O(n^3)$ flops

Build up to it via sequence of simpler algs, also used in practice to find just a few evals/evecs of large/sparse matrices; HQR also used for large sparse matrices (Chapter 7): "Approximate" large sparse matrix by small dense matrix, use HQR on small dense matrix

Plan:

Power Method: Just repeated multiplication of x by A, converges to evec for eval of largest magnitude

Inverse Iteration: Apply power method to $B = (A - \sigma I)^{-1}$ which has same evecs as A, largest eval of B corresponds to eval of A closest to σ.

By choosing σ carefully, can converge to any eval/evec pair

Orthogonal Iteration: ~~Ex~~tends power method from one evec to whole invariant subspace

QR Iteration: Combine Orthogonal Iteration and Inverse Iteration

Other techniques:
- to get down to $O(n^3)$
- real Schur form
- minimize communication
- (discuss some later)

Power Method:

$i=0$, given x_0
repeat
  $y_{i+1} = A x_i$
  $x_{i+1} = y_{i+1} / \|y_{i+1}\|_2$ ... approx evec
  $\lambda'_{i+1} = x_{i+1}^T A x_{i+1}$ ... approx eval
  $i = i+1$
until convergence

Convergence:

$A = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$

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

$x_i = A^i x_0 / \|A^i x_0\|_2$

$= [\lambda_1^i x_0(1), \lambda_2^i x_0(2), \ldots]^T / \| '' \|_2$

$= \lambda_1^i [x_0(1), \left(\frac{\lambda_2}{\lambda_1}\right)^i x_0(2) \ldots]^T / \| '' \|_2$

< 1

as i grows, converges to

$\lambda_1^i [x_0(1), 0 - - - 0]^T / \| '' \|_2$

= evec of λ_1

convergence is $\left|\frac{\lambda_2}{\lambda_1}\right|^i$

Suppose A diagonalizable $A = S \Lambda S^{-1}$

$A^i = S \Lambda^i S^{-1} = S \,\mathrm{diag}(\lambda_1^i \cdots \lambda_n^i) S^{-1}$

$A^i x_0 = S \Lambda^i S^{-1} x_0$ $\qquad z = S^{-1} x_0$

$= S \Lambda^i z$

$= S [\lambda_1^i z_1, \lambda_2^i z_2, \ldots]^T$

$= \lambda_1^i S [z_1, \left(\frac{\lambda_2}{\lambda_1}\right)^i z_2, \ldots]^T$

$\rightarrow \lambda_1^i S [z_1, 0 \cdots 0]^T$

$= \lambda_1^i \cdot z_1 \, S(:, 1) \Rightarrow$ evec of λ_1

To converge at good rate, need

(1) $\left|\frac{\lambda_2}{\lambda_1}\right| < 1$, smaller the better

can't count on this

eg A orthogonal $\Rightarrow$ all $|\lambda_i| = 1$

(2) z_1 nonzero, or if pick x_0 randomly, prob(z_1 small) is small

Inverse Iteration: fix case $|\lambda_1| \approx |\lambda_2|$

power method on $B = (A - \sigma I)^{-1}$

σ called "shift"

$i = 0$, x_0 given

repeat

$y_{i+1} = (A - \sigma I)^{-1} x_i$

$x_{i+1} = y_{i+1} / \|y_{i+1}\|_2$

$\lambda_{i+1}' = x_{i+1}^T A x_{i+1}$

$i = i+1$

until convergence

evecs of B same as for A

evals of B are $1/(\lambda_i - \sigma)$

Suppose σ closest to λ_k

Same analysis as above says

$$\begin{bmatrix} [(\lambda_k-\sigma)/(\lambda_1-\sigma)]^i \frac{z_1}{z_k} \\ [(\lambda_k-\sigma)/(\lambda_2-\sigma)]^i \frac{z_2}{z_k} \\ \vdots \\ 1 \\ \vdots \\ [(\lambda_k-\sigma)/(\lambda_n-\sigma)]^i \frac{z_n}{z_k} \end{bmatrix} \begin{matrix} \\ \\ \\ \leftarrow k^{th} \text{ component} \\ \\ \\ \end{matrix}$$

if σ closer to λ_k than any other λ_i
all factors will be <1, can
converge very fast
where do we get σ? use λ'_{i+1}
Convergence quadratic,
even cubic if $A=A^H$

Next algorithm: converge to
invariant subspace

Orthogonal iteration

given Z_0, $n \times p$ orthogonal matrix
$i=0$
repeat
$Y_{i+1} = A Z_i$
factor $Y_{i+1} = \underset{\text{orthog}}{Z_{i+1}} \cdot \underset{\text{tri}}{R_{i+1}}$ ... QR factorization
... Z_{i+1} spans approximate

invariant subspace

$i = i+1$

until convergence

$p=1$: same as power method

Informal Analysis:

$A = S \Lambda S^{-1}$ diagonalizable

$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_p| > |\lambda_{p+1}| \geq \cdots \geq |\lambda_n|$

need gap for convergence

$\mathrm{span}(Z_{i+1}) = \mathrm{span}(Y_{i+1}) = \mathrm{span}(AZ_i)$

$= \mathrm{span}(A^i Z_0)$ by induction

$= \mathrm{span}(S \Lambda^i S^{-1} Z_0)$

$$A^i Z_0 = S \Lambda^i S^{-1} Z_0$$

$$= S \lambda_p^i \,\mathrm{diag}\Big(\overbrace{\left(\frac{\lambda_1}{\lambda_p}\right)^i, \ldots, \left(\frac{\lambda_{p-1}}{\lambda_p}\right)^i}^{\geq 1}, 1, \overbrace{\left(\frac{\lambda_{p+1}}{\lambda_p}\right)^i \ldots}^{<1}\Big) S^{-1} Z_0$$

$$= S \lambda_p^i \begin{array}{c} p \\ \begin{bmatrix} V_i \\ W_i \end{bmatrix} \end{array} \begin{array}{l} \\ p \\ n-p \end{array}$$

V_i multiplied by $\left(\frac{\lambda_k}{\lambda_p}\right)^i$, $\left(\frac{\lambda_k}{\lambda_p}\right) \geq 1$

getting bigger

W_i multiplied by $\left(\frac{\lambda_k}{\lambda_p}\right)^i$, $\left(\frac{\lambda_k}{\lambda_p}\right) < 1$

getting smaller

i.e. $W_i \to 0$

V_i grows, stays full rank

$A^i Z_0 \to \lambda_p^i S \begin{bmatrix} V_i \\ 0 \end{bmatrix}$ = linear comb. of leading p cols of S

= invariant subspace spanned by first k evecs

First col of Z_i same as power method

First s cols of Z_i same as running with $p = s$

Algorithm computing "top p" invariant subspaces at same time

$\Rightarrow$ Othog. Iter. gives first p invariant subspaces, assuming $|\lambda_1| > |\lambda_2| > \dots$

$\Rightarrow$ Why not let $p = n$, $Z_0 = I$ compute n invariant subspaces. (obstable! real matrices with complex evals have $|\lambda| = |\bar{\lambda}|$)

Thm: Run Othog iter. on A with $Z_0 = I$, $|\lambda_1| > |\lambda_2| > \dots$ and all submatrices $S(1{:}k, 1{:}k)$ have full rank then $A_i = Z_i^T A Z_i$ (similar to A, Z_i orthogonal)

converges to Schur form,

$A_i \to$ upper triangular

evals on diag

proof: for each k, span of first k columns of Z_i converge to invariant subspace spanned by first k evecs of A

$$Z_i = \begin{array}{cc} k & n-k \\ \left[Z_{i1}, \right. & \left. Z_{i2} \right] \end{array} n$$

$$Z_i^H A Z_i = \begin{bmatrix} Z_{i1}^H \\ Z_{i2}^H \end{bmatrix} A \left[Z_{i1}, Z_{i2} \right]$$

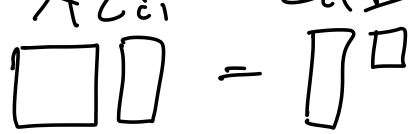
$$= \begin{array}{c|cc} & k & n-k \\ \hline k & Z_{i1}^H A Z_{i1} & Z_{i1}^H A Z_{i2} \\ n-k & Z_{i2}^H A Z_{i1} & Z_{i2}^H A Z_{i2} \end{array}$$

if $\to 0$ for any k

whole matrix $\to$ upper triang.

$Z_{i1} \to$ invariant subspace

$A Z_{i1} \to Z_{i1} B^{k \times k}$



$$Z_{i2}^H A Z_{i1} \to \underbrace{Z_{i2}^H Z_{i1}}_{\to 0 \text{ by orthog of } Z_i} B$$

(see typed notes for code
for Matlab demo of convergence)