Welcome to Ma221! (Mar 15)

Perturbation Theory
Can I trust my answer?

Last time: $A = I$ showed eigenvectors very sensitive: takeaway. Close evs $\Rightarrow$ evcs ill conditioned

To describe perturbations in evs:

Def: Epsilon-pseudo-spectrum of $A$

is set of all evs of all matrices within distance $\varepsilon$ of $A$:

$$\Lambda_\varepsilon (A) = \{ \lambda : (A+E)x = \lambda x \text{ for some } x \neq 0 \} \text{ s.t. } \|E\|_2 \leq \varepsilon$$

Smallest possible $\Lambda_\varepsilon (A)$:

- Each disk around eval of $A$, radius $\varepsilon$
  - attained by $E = \varepsilon I$
  - attained by $A = A^\dagger$ (Chap 5)

Worst (most sensitive) case

Thus (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 \( J(n,1) = \varepsilon \begin{bmatrix} 0 & 1 \\ \varepsilon & 0 \end{bmatrix} \)

\( \rho(A) = \lambda^2 - \varepsilon \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \)

\( \Rightarrow \lambda = \sqrt{\varepsilon} \) uniformly spaced on circle of radius \( \sqrt{\varepsilon} \)

\( \varepsilon = 10^{-16} \), \( n = 16 \)

\( \sqrt{\varepsilon} = 0.1 \)

(1) evals are continuous but not necessarily differentiable

(slope of \( \sqrt{\varepsilon} = \infty \) at \( \varepsilon = 0 \))

(2) expect sensitive evals when evals nearly multiple
Condition number of simple (nonmultiple) evals (else $\infty$)

Thm: $\lambda$ be simple eval of $A$

$Ax = \lambda x$, $y^*Ax = \lambda y^*x$, $\|x\|_2 = \|y^*x\|_2 = 1$

If we perturb $A$ to $A + E$

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

$\delta \lambda = \frac{y^*Ex}{y^*x} + O(\|E\|_2)$

$|\delta \lambda| \leq \frac{\|E\|_2}{\|y^*x\|} + O(\|E\|_2)$

$= \sec(\Theta) \cdot \|E\|_2 + O(\|E\|_2)$

$\Theta = \text{angle between } x \text{ and } y$

$\sec(\Theta) = \text{condition number}$

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

$Ax + A \delta x + Ex + E \delta x = \lambda x + \lambda \delta x + \delta \lambda x + \delta \lambda \delta x$

cancel $Ax$ second order, drop $\delta \lambda \delta x$

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

$y^*A \delta x + y^*Ex = y^* \lambda \delta x + y^* \delta \lambda x$

cancel

$\frac{y^*Ex}{y^*x} = \delta \lambda$
Special Case 1: \( A = A^H \) or "normal"
\[ AA^H = A^H A \] (HWQ 4.2)
\[ \implies \text{A has orthonormal evecs} \]

Cor: If A normal, perturbing A to \( A + \delta \)
\[ 1/|\delta| \leq 1/\|E\|_2 + O(\|E\|_2^2) \]
\[ \text{i.e. condition } \# = 1 \]

Proof: \( A = Q \Lambda Q^T \)
\[ AQ = Q \Lambda \text{ and } Q^T A = \Lambda Q^T \]
right evecs = cols of \( Q \)
left evecs = rows of \( Q^T = \text{cols of } Q \)
\[ \implies x = y \implies y^H x = 1 \]

Later (Chap 5) if \( A = A^H \) and \( E = E^H \)
then \[ 1/|\delta| \leq 1/\|E\|_2 \] (no \( 1/\|E\|_2^2 \) term)

Special Case 2: \( A = \text{Jordan Block} \)
\[ A = \begin{bmatrix} 0 & 1 \\ \vdots & \ddots & \ddots \\ 0 & \cdots & 0 & 1 \\ 0 & \cdots & 0 & 0 & 1 \end{bmatrix} \]
\[ x = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \]
\[ y = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \]
\[ y^H x = 0 \implies \text{cond } \# = \infty \]

Extend to eliminate \( O(\|E\|_2^2) \) term
Thm (Bauer–Fike) if $A$ has all simple evals, $\lambda_i$, with right and left evcs $x_i$ 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 $\lambda_i$
with radii $\frac{\|E\|_2}{\|y_i^* x_i\|}$

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

Algorithms for Non-Symmetric Eigenproblem

Ultimate Algorithm:
Hessenberg QR (HQR)
Takes any $n \times n$ nonsymmetric dense $A$
computes Schur form $A = Q^* T Q$
in $O(n^3)$ flops
Build up to it via sequence of simpler algs, also used in practice to find just a few evals/evects 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 evect for eval of largest magnitude

Inverse Iteration: Apply power method to \( B = (A - \sigma I)^{-1} \) which has same evects as A, largest eval of B corresponds to eval of A closest to \( \sigma \). By choosing \( \sigma \) carefully, can converge to any eval/evect pair.
Orthogonal Iteration: Extends power method from one eigenvector 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:
- $i = 0, \text{ given } \mathbf{x}_0$
- repeat
  - $y_{i+1} = A \mathbf{x}_i$
  - $\mathbf{x}_{i+1} = y_{i+1} / \| y_{i+1} \|_2$ ... approx. eigenvector
  - $\lambda_{i+1} = \mathbf{x}_{i+1}^T A \mathbf{x}_{i+1}$ ... approx. eigenvalue
- $i = i + 1$
- until convergence.
Convergence:

\[ A = \text{diag} \left( d_1, d_2, \ldots, d_n \right) \]
where \( |d_1| > |d_2| \geq \ldots \geq |d_n| \)

\[ x_i = A^i x_0 / \| A^i x_0 \|_2 \]
\[ = \left[ \lambda_1^i x_0(1), \lambda_2^i x_0(2), \ldots \right]^T / \| \cdot \|_2 \]
\[ = \lambda_1^i \left[ x_0(1), \left( \frac{d_2}{d_1} \right)^i x_0(2), \ldots \right]^T / \| \cdot \|_2 \]
\[ \rightarrow \lambda_1 \]

as \( i \) grows, converges to

\[ \lambda_1 \left[ x_0(1), 0, \ldots, 0 \right]^T / \| \cdot \|_2 \]
\[ = \text{evec of } \lambda_1 \]

cconvergence is \( \left| \frac{d_2}{d_1} \right|^i \)

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

\[ A^i = S \Lambda^i S^{-1} = S \text{diag} \left( \lambda_1^i, \ldots, \lambda_n^i \right) S^{-1} \]

\[ A^i x_0 = S \Lambda^i S^{-1} x_0 \]
\[ = S \Lambda^i z \]
\[ = S \left[ \lambda_1^i z_1, \lambda_2^i z_2, \ldots \right]^T \]
\[ = \lambda_1^i S \left[ z_1, \left( \frac{d_2}{d_1} \right)^i z_2, \ldots \right]^T \]
\[ \rightarrow \lambda_1 \left[ z_1, 0, \ldots, 0 \right]^T \]
\[ = \lambda_1^i z_1 \text{ } S(:, 1) \Rightarrow \text{evec of } \lambda_1 \]
To converge at good rate, need
(1) \( \left| \frac{d\lambda}{d\xi} \right| < 1 \), smaller the better
can't count on this
e.g. \( A \) orthogonal \( \Rightarrow \lambda_i \left| \xi_i \right| = 1 \)

(2) \( \xi_i \) nonzero, or if pick \( \xi_0 \) randomly, \( \text{prob}(\xi_i \text{ small}) \) is small

Inverse Iteration: fix case \( 1 \leq |\xi| \leq 1 \)

Power method on \( B = (A - \sigma I)^{-1} \)
called "shift"

\( i = 0 \), \( x_0 \) given

Repeat
\( y_{i+1} = (A - \sigma I)^{-1} x_i \)
\( \xi_{i+1} = y_{i+1} / \| y_{i+1} \|_2 \)
\( \lambda_{i+1} = \xi_{i+1}^T A \xi_{i+1} \)
\( \tilde{\xi} = \xi_{i+1} \)

until convergence

Evecs of \( B \) same as for \( A \)
Evrs of \( B \) are \( 1/(\tilde{\xi}_i - \sigma) \)
Suppose \( \sigma \) closest to \( \tilde{\xi} \)

Same analysis as above says
\[
\begin{bmatrix}
\frac{(\lambda_k - \sigma)}{(\lambda_i - \sigma)} & \frac{z_i}{\epsilon_k} \\
\frac{(\lambda_k - \sigma)}{(\lambda_i - \sigma)} & \frac{z_i}{\epsilon_i} \\
\vdots & \vdots \\
\frac{(\lambda_k - \sigma)}{(\lambda_i - \sigma)} & \frac{z_i}{\epsilon_k}
\end{bmatrix}^{k_{th}} \text{ component}
\]

If \( \sigma \) closer to \( \lambda_k \) than any other \( \lambda_i \) all factors will be \( < 1 \), can converge very fast.

Where do we put \( \sigma \)? Use it!

Convergence quadratic, even cubic if \( A = A^T \)

Next algorithm: converge to invariant subspace

Orthogonal iteration

Given \( Z_0 \), n x p orthogonal matrix

\( i = 0 \)

Repeat

\( Y_{i+1} = AZ_i \) \( \text{orthog tri factor} \)

\( Y_{i+1} = Z_{i+1} \cdot R_{i+1} \ldots QR \)

\( Z_{i+1} \) spans approximate
Invariant subspace

\[ \mathcal{v} = \mathcal{v} + 1 \]

until convergence

\[ p = 1 : \text{same as power method} \]

**Informal Analysis:**

\[ A = S \Lambda S^{-1} \text{ diagonalizable} \]

\[ 1 \lambda_1 \geq 1 \lambda_2 \geq \ldots \geq 1 \lambda_p > 1 \lambda_{p+1} \geq \ldots \geq 1 \lambda_n \]

\[ \text{need gap for convergence} \]

\[ \text{span}(Z \mathcal{v} + 1) = \text{span}(\mathcal{v} + 1) = \text{span}(AZ) \]

\[ = \text{span}(A^2 Z) \text{ by induction} \]

\[ = \text{span}(S \Lambda \mathcal{v} S^{-1} Z) \]

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

\[ = S \Lambda^i \text{diag}(\frac{d_1}{\lambda_p}, \ldots, \frac{d_p}{\lambda_p}, 1, \frac{d_{p+1}}{\lambda_p}, \ldots) S^{-1} Z_0 \]

\[ = S \Lambda^i \left[ \begin{array}{c} \mathcal{v} \\ W_i \end{array} \right] n-p \]

\[ \mathcal{v} \text{ multiplied by } \left( \frac{d_k}{\lambda_p} \right)^i, \left( \frac{\lambda_k}{\lambda_p} \right)^i \geq 1 \]

\[ \text{getting bigger} \]

\[ W_i \text{ multiplied by } \left( \frac{d_k}{\lambda_p} \right)^i, \left( \frac{n_k d_k}{\lambda_p} \right) \leq 1 \]

\[ \text{getting smaller} \]

i.e. \[ W_i \rightarrow 0 \]
Vi grows, stays full rank
\[ A^t Z_0 \rightarrow A^t \begin{bmatrix} V_1 & \cdots & V_p \end{bmatrix} = \text{linear comb. of leading } p \text{ cols of } S \]

= invariant subspace spanned by first k evec

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

\( \Rightarrow \) Orthog. Iter. gives first \( p \) invariant subspaces, assuming \( |1| > |1_2| > \cdots \)

\( \Rightarrow \) Why not let \( p = n \), \( Z_0 = I \)
compute \( n \) invariant subspaces.

(obstacle: real matrices with complex evals have \( |1| = |1_2| \))

Thm: Run Orthog iter. on \( A \) with
\( Z_0 = I \), \( |1| > |1_2| > \cdots \)
and all submatrices \( S(1:k, 1:k) \) have full rank
Then $A_i = Z_i^T A Z_i$ (similar to $A$ with $Z_i$ orthogonal) converges to Schur form $A \rightarrow upper\; triangular\; forms\; on\; diag$.

Proof: For each $k$, span of first $k$ columns of $Z_i$ converge to invariant subspace spanned by first $k$ eigenvalues of $A$.

$Z_i = \begin{bmatrix} Z_{i1}^T, Z_{i2}^T \end{bmatrix}$

$Z_i^T A Z_i = \begin{bmatrix} Z_{i1}^T \ A \ Z_{i1}^T & Z_{i1}^T \ A \ Z_{i2}^T \\ Z_{i2}^T \ A \ Z_{i1} & Z_{i2}^T \ A \ Z_{i2} \end{bmatrix}$

$= \begin{bmatrix} \ Z_{i1}^T \ A \ Z_{i1} & Z_{i1}^T \ A \ Z_{i2} \\ Z_{i2}^T \ A \ Z_{i1} & Z_{i2}^T \ A \ Z_{i2} \end{bmatrix}$

$Z_{i1} \rightarrow invariant\; subspace$

$A Z_{i1} \rightarrow Z_{i1} B^{k \times k}$
$Z_i^h A Z_i^h \rightarrow Z_i^h Z_i^h B$

$\rightarrow 0$ by orthog of $Z_i$

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