Welcome to Ma221! Lecture 17, Fall 2024

Chap 4: Nonsymmetric Eigenproblem

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

$i = 0$, 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:

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

where $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \dots$

$\uparrow$ necessary for convergence

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

$$= [\lambda_1^i x_0(1), \lambda_2^i \cdot x_0(2), \dots]^T / \|\,"\,\|$$

$$= \lambda_1^i [x_0(1), \underbrace{\left(\frac{\lambda_2}{\lambda_1}\right)}_{<1}^i x_0(2), \dots]^T / \|\,"\,\|$$

$\longrightarrow \lambda_1^i [x_0(1), 0, 0 \cdots 0]^T / \| " \|_2$

$\longrightarrow [1, 0, \ldots \ldots 0] =$ evec for λ_1

convergence depends on $\left(\frac{\lambda_2}{\lambda_1}\right)^i$

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

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

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

$$= S \Lambda^i z$$

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

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

as $i \to \infty$ $\longrightarrow \lambda_1^i \cdot S [z_1, 0, \ldots 0]^T$

$= \lambda_1^i z_1 S(:,1) =$ first col of S

$=$ evec for λ_1

To converge at good rate: need

① $\left|\frac{\lambda_2}{\lambda_1}\right| < 1$ smaller is better, not always true

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

② z_1 nonzero, not too small,

ok to pick x_0 randomly,

prob(z_1) small is small

Inverse Iteration: fix case $\left|\frac{\lambda_2}{\lambda_1}\right|$ near 1

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

σ = "shift"

$i = 0$, x_0 given

repeat

$y_{i+1} = (A-\sigma I)^{-1} x_0$

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

$\lambda_{i+1}' = x_{i+1}^T A x_{i+1}$ ← source of σ

$i = i+1$

until convergence

evecs of B same as for A

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

suppose σ closest to λ_k

Same analysis as above:

$$\begin{bmatrix} ((\lambda_k-\sigma)/(\lambda_1-\sigma))^i \\ ((\lambda_k-\sigma)/(\lambda_2-\sigma))^i \\ \vdots \\ 1 \\ \vdots \\ ((\lambda_k-\sigma)/(\lambda_n-\sigma))^i \end{bmatrix} \quad k^{th} \text{ component}$$

if σ closer to λ_k than any other λ_i

all these factors $\left|\frac{\lambda_k-\sigma}{\lambda_i-\sigma}\right| << 1$

$\Rightarrow$ converge fast to $\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ as $i \to \infty$

Where do we get σ? use λ_{i+1}

$\Rightarrow$ Quadratic convergence

even cubic convergence if $A = A^T$

Next algorithm: converge to whole invariant subspace

Orthogonal Iteration:

given Z_0, $n \times p$ orthogonal matrix

$i = 0$

repeat

$Y_{i+1} = A Z_i$

factor $Y_{i+1} = \underset{\text{orthog}}{Z_{i+1}} \cdot \underset{\text{triangular}}{R_{i+1}}$ ... QR decomp

... Z_{i+1} spans approximate invariant subspace

until convergence

$p = 1$: same as power method

Informal Analysis:

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

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

needed for convergence

$$\text{span}(Z_{i+1}) = \text{span}(Y_{i+1}) = \text{span}(A Z_i)$$
$$= \text{span}(A^i Z_0)$$
$$= \text{span}(S \Lambda^i S^{-1} Z_0)$$

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

$$= S \lambda_p^i \,\mathrm{diag}\Big(\underbrace{\left(\tfrac{\lambda_1}{\lambda_p}\right)^i, \dots, \left(\tfrac{\lambda_{p-1}}{\lambda_p}\right)^i, 1}_{\geq 1}, \overbrace{\left(\tfrac{\lambda_{p+1}}{\lambda_p}\right)^i, \dots}^{<1} \Big) \cdot S^{-1} Z_0$$

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

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

getting bigger, stays full rank

W_i multiplied by $\left(\frac{\lambda_k}{\lambda_p}\right)^i < 1 \Rightarrow W_i \to 0$

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

= invariant subspace of leading p evecs

First col of Z_i: same as power method

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

Algorithm computes "top p" invariant subspaces at same time, assuming $|\lambda_1| > |\lambda_2| > \dots$

$\Rightarrow$ Why not pick $p = n$ with $Z_0 = I$?

compute n invariant subspaces

(assuming $|\lambda_1| > |\lambda_2| > \dots$ for now)

Thm: Run Orthog Iteration on A with $Z_0 = I$, assume $|\lambda_1| > |\lambda_2| > \dots$ and all Submatrices $S(1:k, 1:k)$ have full rank

then $A_i = Z_i^H A Z_i$ (A_i similar to A)
converges to Schur form
$A_i \to$ upper triangular with
$\lambda_1, \lambda_2, \ldots$ on diagonal

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

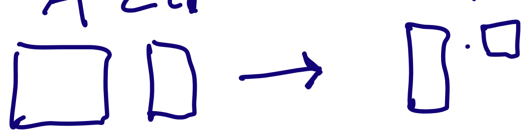
$$Z_i = \begin{array}{cc} \begin{array}{cc} k & n-k \end{array} & \\ \left[\begin{array}{c|c} Z_{i1} & Z_{i2} \end{array}\right] & n \end{array}$$

$$Z_i^H A Z_i = \begin{bmatrix} Z_{i1}^H \\ Z_{i2}^H \end{bmatrix} A \left[\begin{array}{c|c} Z_{i1} & Z_{i2} \end{array}\right]$$

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

if this $\to 0$ for all k
$\Rightarrow$ whole matrix $\to$ upper triangular
= Schur form

$Z_{i1} \to$ invariant subspace
$\Rightarrow A Z_{i1} \to Z_{i1} \cdot B$



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

(see typed notes for matlab demo code)

Next steps: power method → orthogonal iteration → QR iteration → add inverse iteration

lets us converge to any eval for which we have an approximation, to be supplied by the algorithm

QR Iteration

Given $A_0 = A$

$i = 0$

repeat

  factor $A_i = Q_i R_i$

  $A_{i+1} = R_i Q_i$

  $i = i+1$

until convergence

$$A_{i+1} = R_i Q_i = \underbrace{\overbrace{Q_i^T Q_i}^{=I} R_i}_{=A_i} Q_i = Q_i^T A_i Q_i$$

Thm: A_i from QR Iteration same as $Z_i^T A Z_i$ from Orthog Iteration ($Z_0 = I$, $A Z_i = Z_{i+1} R_{i+1}$ ... QR decomp)

$\Rightarrow A_i$ converges to Schur form if $|\lambda_1| > |\lambda_2| > \ldots$

proof: induction on i: assume $A_i = Z_i^T A Z_i$

One Step of Orthog Iteration

$A Z_i = Z_{i+1} R_{i+1}$ ... QR decomp

$$A_i = Z_i^T \underbrace{A Z_i}$$ ... induction assumption

$$= \underbrace{Z_i^T Z_{i+1}}_{\text{orthog}} \underbrace{R_{i+1}}_{\text{triangular}}$$

= QR decomp of A_i by uniqueness of QR

$$Z_{i+1}^T A Z_{i+1} = \underbrace{Z_{i+1}^T A (Z_i}\underbrace{Z_i^T) Z_{i+1}}$$

$$= R_{i+1} \cdot (Z_i^T \cdot Z_{i+1})$$

$$= R \cdot Q$$

$= A_{i+1}$ from QR QED of induction

Add inverse iteration: QR iteration with shift

$A_0 = A$

$i = 0$

repeat

choose shift σ_i near eval

factor $A_i - \sigma_i I = Q_i R_i$

$A_{i+1} = R_i Q_i + \sigma_i I$

$i = i+1$

until convergence

Lemma: A_i and A_{i+1} are orthog. simila

pf. $A_{i+1} = R_i Q_i + \sigma_i I$

$$= Q_i^T \underbrace{Q_i R_i} Q_i + \sigma_i I$$

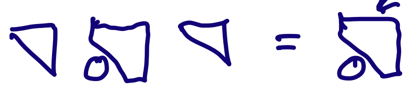
$$= Q_i^T (A_i - \sigma_i I) Q_i + \sigma_i I$$

$$= Q_i^T A_i Q_i \underbrace{- \sigma_i Q_i^T Q_i + \sigma_i I}_{\text{cancel}}$$

$= Q_i^T A_i Q_i$ similar QED

Note: if R_i nonsingular, can also write

$$A_{i+1} = R_i Q_i + \sigma I$$
$$= R_i \underbrace{Q_i R_i} R_i^{-1} + \sigma_i I$$
$$= R_i (A_i - \sigma_i I) R_i^{-1} + \sigma_i I$$
$$= R_i A_i R_i^{-1}$$

=  ← upper Hessenberg

use this later to get cost of 1 iteration down to $O(n^2)$ from $O(n^3)$

If σ_i exact eval of A,

QR with shift σ_i converges in one step

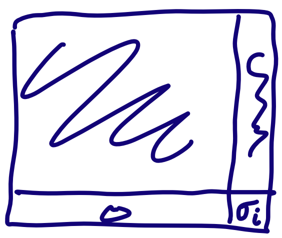
$A_i - \sigma_i I$ singular $\Rightarrow R_i$ singular

$\Rightarrow R_i(k,k) = 0$, suppose $R_i(n,n) = 0$

$\Rightarrow$ last row of $R_i = 0$

$\Rightarrow$ last row of $R_i Q_i = 0$

$\Rightarrow$ last row of $A_{i+1} = R_i Q_i + \sigma_i I$ is zero except $A_{i+1}(n,n) = \sigma_i$

$A_{i+1} =$ 

block triangular → evals are evals of diagonal blocks including σ_i

algorithm proceeds, working only on $A_{i+1}(1:n-1, 1:n-1)$

If σ_i not exact eval, declare convergence if

$\|A_{i+1}(n, 1:n-1)\| = O(\varepsilon) \cdot \|A\|$,

then set $A_{i+1}(n, 1:n-1) = 0$, backward stable

(see typed notes for code for matlab demo)

Previous analysis: expect

$A_{i+1}(n, 1:n-1)$ to shrink by factor

$$\frac{|\lambda_k - \sigma_i|}{\min_{j \neq k} |\lambda_j - \sigma_i|} < 1$$

(λ_k = closest eval to shift σ_i)

[Finish proof of quadratic convergence next time]