Eigenvalue problems

① Recap from LA

Def. Matrix $A \in \mathbb{R}^{n \times n}$

Eigenvalue/-vector: $Ax = \lambda x$ for vector $x \neq 0$

eigenvector uniquely det.?

Spectrum: $\lambda(A)$: set of eigenvalues

Spectral radius: $\rho(A) = \max\{|\lambda| : \lambda \in \lambda(A)\}$

Finding:

λ eigenvalue $\Leftrightarrow (A - \lambda I)x = 0$ solvable with $x \neq 0$

$\Leftrightarrow A - \lambda I$ singular

$\Leftrightarrow \det(A - \lambda I) = 0$

char. polynomial

degree n

has n roots (possibly in $\mathbb{C}$)

Abel (1824): cannot write down formula for roots of polynomial w/ deg ≥ 5.

→ algorithmic consequence?

0 left / right eigenvector
left / right eigenvalues?

Multiplicity:

Algebraic $\longleftrightarrow$ Geometric

multiplicity as root $\geq$ # lin. indep. eigenvec.

$>$ $\rightarrow$ A "defective"

Example: $\begin{pmatrix} 1 & 1 \\ & 1 \end{pmatrix}$

CP?

eigenvalues?

eigenvectors?

$$\begin{pmatrix} 1 & 1 \\ & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \rightsquigarrow x + y = x$$

A not defective: Diagonalizable

$$A = XDX^{-1}$$

$\begin{pmatrix} \diagdown \end{pmatrix}$ diagonal

similarity transform

with $X = \begin{bmatrix} x_1 & \cdots & x_n \\ & & \\ & & \end{bmatrix}$ eigenvectors

② Sensitivity

Assume A not defective

Suppose $X^{-1}AX = D$

Perturb $A \longrightarrow A+E$: What happens to eigenvalues?

$X^{-1}(A+E)X = D+F$ ⟵ $(A+E)$ and $(D+F)$ have same λ

not necessarily diagonal

Suppose v is a perturbed eigenvector:

$$(D+F)v = \mu v$$

$$\Leftrightarrow \quad Fv = (\mu I - D)v$$

$$\Leftrightarrow \quad (\mu I - D)^{-1} F v = v$$

when invertible?

$$\Rightarrow \quad \|(\mu I - D)^{-1}\|_2^{-1} \leq \|F\|_2$$

diagonal, so = ?

$$|\mu - \underbrace{\lambda_k}_{\text{closest to } \mu}| = \|(\mu I - D)^{-1}\|_2^{-1}$$

$$\leq \|F\|_2 = \|X^{-1}EX\|_2$$

$$\leq \text{cond}_2(X)\, \|E\|_2$$

Meaning?

Bad situation in X?

Meaning for individual eigenvalue?

How about a single eigenvalue?

LEFT OUT

Let $Ax = \lambda x$, $y^T A = \lambda y^T$.

multipl.=1 $\Rightarrow y^T x \neq 0$

$$(A+E)(x+\Delta x) = (\lambda + \Delta\lambda)(x+\Delta x)$$

drop $\Delta(\text{something})^2$

$$\underbrace{\cancel{Ax}}_{\lambda x} + A\Delta x + Ex \approx \cancel{\lambda x} + \lambda \Delta x + \Delta\lambda x$$

$$\underbrace{y^T A}_{\lambda y^T}\Delta x + y^T E x \approx \lambda y^T \Delta x + \Delta\lambda \, y^T x$$

$$y^T E x \approx \Delta\lambda \; y^T x$$

$$\Delta\lambda \approx \frac{y^T E x}{y^T x}$$

$$\leadsto \quad |\Delta\lambda| \leq \frac{\|y\| \cdot \|E\| \, \|x\|}{|y \cdot x|} \leq \frac{1}{\cos\angle xy} \|E\|$$

Meaning?

Symmetric mat $\Rightarrow x = y \rightarrow$?

<Lec 13>

Announcements:

- lecture material shuffle
- Health slides

Recap:

$A \to A + E$

$|\mu - \lambda_k| \leq \mathrm{cond}_2(X)\, \|E\|_2$

eigenvalue of A closest to μ

3 Properties and Transformations

Suppose $Ax = \lambda x$.

Shift $A - \sigma I$ $\qquad (A - \sigma I)x = (\lambda - \sigma)x$

Inversion A^{-1} $\qquad A^{-1}x = \lambda^{-1}x$

Power A^k $\qquad A^k x = \lambda^k x$

Polynomial $aA^2 + bA + cI$ $\qquad (aA^2 + bA + c)x = (a\lambda^2 + b\lambda + c)x$

Similarity $T^{-1}AT$ $\qquad y := T^{-1}x$

non-singular

$$T^{-1}ATy = \lambda y$$

eigenvalues?

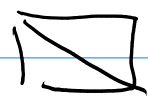
Orthonormal: $A^TA=I$ $\quad$ $A^HA=I$ (unitary)

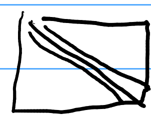
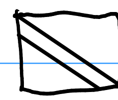
Symmetric: $A=A^T$ $\quad$ $A=A^H$ (hermitian)

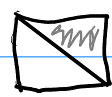
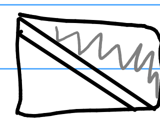
real eigenvalues

Normal $\quad$ $A^HA=AA^H$

orth. eigenvectors

Diagonal:  ✓

Tridiagonal:  $\quad$ Bidiagonal 

Triangular: "Schur"  ✓ $\quad$ Hessenberg 

✓ = eigenvalues on diagonal

Tool: Schur factorization

Every matrix is orthonormally similar to an upper triangular matrix.

$$A = QUQ^T$$

If we knew how to compute this, why would it be helpful for knowing eigenvalues?

Existence? (assume non-defective)

Suppose $Av = \lambda v$. $(\neq 0)$
Let $V = \operatorname{span}\{v\}$.

Then: $A: V \longrightarrow V$

$\qquad V^\perp \longrightarrow V \oplus V^\perp$

Write that in matrix form:

$$A = \underbrace{\begin{pmatrix} v & \begin{matrix}\text{Basis of}\\ V^\perp\end{matrix} \end{pmatrix}}_{Q_1} \begin{array}{c} \begin{pmatrix} \lambda & \boxed{\quad * \quad} \\ \underbrace{\boxed{0}}_{v} & * \end{pmatrix} \end{array} \begin{array}{l} \}v \\ \\ \end{array} Q_1^T$$

Repeat: Matrix becomes triangular.

What happens w/ complex λ?
w/ defective A?

WS 12 p1

④ Computing

Power iteration

Eigenvalues of $A^{20,000}$?

$|\lambda_1| > |\lambda_2| > \ldots$

$\uparrow \quad \uparrow$

$x_1 \quad x_2$

Assume $\|x_i\| = 1$

$$y = A^{20,000}(\underbrace{\alpha x_1 + \beta x_2}_{x}) = \alpha \lambda_1^{20,000} x_1 + \beta \lambda_2^{20,000} x_2$$

$$\frac{y}{\lambda_1^{20,000}} = \alpha x + \underbrace{\beta \Big(\underbrace{\frac{\lambda_2}{\lambda_1}}_{<1}\Big)^{20,000}}_{\ll 1}$$

Possible problems:

- starting vector has no component for λ_1
- $\lambda_1 = \lambda_2$

Computational risk? (Overflow)

→ "normalized" power it.

<lec 14>

WS 13 pb 1

Convergence of Power iteration

$e_n = \| \lambda_1 - \hat{\lambda}_1^{(k)} \|$ $\qquad e_{n+1} \approx \frac{|\lambda_2|}{|\lambda_1|} \cdot e_n$ ← slow

Shift can help:

$$e_{n+1} \approx \frac{|\lambda_2 - \sigma|}{|\lambda_1 - \sigma|} \cdot e_n$$

When is this small? → not easy to say!

Don't even know λ_1! But: can estimate!

Rayleigh quotient: given approx. eigenvector, est. eigenvalue

$$\frac{x^T A x}{x^T x} \approx \lambda$$

Inverse iteration: Power iteration with A^{-1}

NOTE: Shift + Inversion can make any eigenvalue the largest in magnitude.
→ can compute any eigenvector (at least in theory?)

Convergence for inverse it:

$$e_{n+1} \approx \frac{|\lambda_{\text{closest to } \sigma} - \sigma|}{|\lambda_{\text{second-closest to } \sigma} - \sigma|} \cdot e_n$$

→ shifting by eval belonging to conv. target actually good!

→ use Rayleigh quotient

Demo: Power Iteration

WS 13 pb 2

Downside: one at a time

- Idea 1: peel off an eigenvalue, start again
  ↳ "Deflation"
  = similarity transf. to $\begin{pmatrix} \lambda_1 & * \\ & \boxed{B} \end{pmatrix}$

  [(One step in constr. of Schur form)

- Idea 2: iterate with multiple vectors

Simultaneous Iteration:

$X_0 \in \mathbb{R}^{n\times p}$ arbitrary $(p \leq n)$

$X_{k+1} = AX_k$

[Problems? → needs rescaling
→ increasingly ill-conditioned (even post-rescale)

Orthogonal Iteration

$X_0 \in \mathbb{R}^{n\times p}$ arbitrary $(p\leq n)$

$Q_k R_k = X_k$ (reduced)

$X_{k+1} = AQ_k$

(X_k) converges to X with $X = \begin{bmatrix} x_1 & \cdots & x_p \end{bmatrix}$ $\quad Ax_i = \lambda_i x_i$
$\quad |\lambda_i| > |\lambda_{i+1}|$

[Problems? → slow/linear convergence
→ expensive iteration

What's it do?

(assume X_0 has full rank)

$$Q_0 R_0 = X_0$$
$$X_1 = AQ_0$$
$$Q_1 R_1 = X_1 \quad = AQ_0 \Rightarrow Q_1 R_1 Q_0^T = A$$
$$X_2 = AQ_1$$
$$Q_2 R_2 = X_2 \quad = AQ_1 \Rightarrow Q_2 R_2 Q_1^T = A$$

Once Q converges ($\Rightarrow Q_{n+1} \approx Q_n$) have Schur factorization!

Q: What's $Q_3 R_3 Q_3^T$?

Once the Q's converge, this becomes A, too!

What's $Q_3^T A Q_3 =: \hat{X}_3$

Approximately R_3 → conv. to upper triangular.

Idea: Magnitude of below-diag part of $\hat{X}_3$ ought to help monitor convergence.

Demo: Convergence to Schur form

<Lec 15>

Saw: Orth. iteration produces (mostly) conv. sequence of (Q_i)

$$Q_i R_i Q^T_{i-1} = A$$

At convergence, $Q_i = Q_{i+1} \rightarrow$ Schur form.

Better conv. indicator: $Q_i^T A Q_i =: \hat{X}_i$

(conv. to lower triang.)

Let's try to compute $\hat{X}_k$ directly.

Suppose we started with $X_0 = A$.

"old"

$X_0 = A$

1 $Q_0 R_0 = X_0$

2 $\hat{X}_0 = Q_0^T A Q_0$

$= Q_0^T (Q_0 R_0) Q_0$

$= R_0 Q_0$

4 $X_1 = A Q_0$

$= Q_0 Q_0^T A Q_0$

$= Q_0 \bar{X}_1$

$= Q_0 (\bar{Q}_1 \bar{R}_1)$

$= (Q_0 \bar{Q}_1) \bar{R}_1$

$Q_1 R_1 = X_1$

same

up to sign

"new"

$\bar{X}_0 = A$

$\bar{Q}_0 \bar{R}_0 = \bar{X}_0$

$\bar{X}_1 = \bar{R}_0 \bar{Q}_0 \quad (= \hat{X}_0)$ 3

$\bar{Q}_1 \bar{R}_1 = \bar{X}_1$

$\bar{X}_2 = \bar{R}_1 \bar{Q}_1 \quad = \bar{Q}_1^T \bar{X}_1 \bar{Q}_1$

$= \bar{Q}_1^T \hat{X}_0 \bar{Q}_1$

$= \bar{Q}_1^T Q_0^T A Q_0 \bar{Q}_1^T$

$= Q_1^T A Q_1$

$= \hat{X}_1$

$Q_0 = \bar{Q}_0$

$Q_1 = \bar{Q}_0 \bar{Q}_1$

$Q_k = \bar{Q}_0 \bar{Q}_1 \cdots \bar{Q}_k$

<u>QR Iteration</u>

$\bar{X}_0 = A$

$\bar{Q}_k \bar{R}_k = \bar{X}_k$

$\bar{X}_{k+1} = \bar{R}_k \bar{Q}_k$

We know the $\hat{X}_k$ ($= \bar{X}_{k+1}$) converge (from understanding orthogonal it.)

$\Rightarrow$ QR it. eventually produces Schur form!

WS14 pb1

Still has issues from earlier: $\nearrow$ slow conv. $\searrow$ expensive it.

O Earlier fix for slow convergence?

<u>QR it. with Shift</u>

$\bar{X}_0 = A$

Choose σ_k

$\bar{Q}_k \bar{R}_k = \bar{X}_k - \sigma_k I$

$\bar{X}_{k+1} = \bar{R}_k \bar{Q}_k + \sigma_k I$

$\bar{Q}_0 \bar{R}_0 = A - \sigma_0 I$

$\bar{X}_1 = \bar{R}_0 \bar{Q}_0 + \sigma I$

$= \bar{Q}_0^T (A - \sigma I) \bar{Q}_0^T + \sigma I$

$= \bar{Q}_0^T A \bar{Q}_0 - \sigma \bar{Q}_0^T \bar{Q}_0 + \sigma I$

O So: does shifted QR it. produce the same iterates as regular QR it.?

No, different $\bar{Q}_0$!

- Should choose shift close to existing eigenvalue

LEFT OUT

- On 'exact hit', $\bar{X}_k - \sigma_k I$ becomes singular.
  - → last col/row of $\bar{X}_{k+1}$ zero except for diag
  - → convergence crit
  - → once converged, stop worrying about this row.

O Heuristics for σ_k:

- $(\bar{X})_{nn}$
- eigenvalues of lower right 2×2 of $(\bar{X})_{nn}$

Fix for too much work: Produce zeros

$$\text{Full} \xrightarrow[\text{Householder}]{} \text{Hessenberg} \xrightarrow[\text{QR it.}]{} \text{diagonal}$$

QR it. ⟵ using which QR?

$$\text{Symmetric} \xrightarrow[\text{HH}]{} \text{tridiag.} \xrightarrow[\text{QR it.}]{} \text{diag.}$$

4.5 Krylov space methods

[Have to build subspace in which to search for eigenvectors

QR it: $\quad \operatorname{span}\{A^{\ell} y_1, A^{\ell} y_2, \ldots, A^{\ell} y_k\}$

Krylov: $\quad \operatorname{span}\{\underbrace{x}_{x_0}, \underbrace{Ax}_{x_1}, \ldots, \underbrace{A^{k-1}x}_{x_{k-1}}\}$

Define matrix $K_k := [x_0 \ldots x_{k-1}]$. $\leftarrow n \times k$

$$AK_n = [x_1 \ldots x_{n-1} x_n]$$

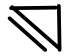
$$= K_n \underbrace{[e_2 \ldots e_n \; K_n^{-1} x_n]}_{C_n :=}$$

$$K_n^{-1} A K_n = C_n = \begin{bmatrix} \diagdown & 0 & \vert \\ 0 & \diagdown & \vert \end{bmatrix} \leftarrow \text{upper Hessenberg}$$

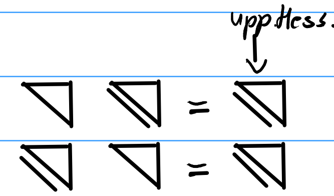
[Conditioning of K_n? Fix?

$Q_n R_n = K_n \quad \to \quad Q_n = K_n R_n^{-1}$

$$Q_n^T A_n Q_n = Q_n^{-1} A_n Q_n = R_n \underbrace{K_n^{-1} A K_n}_{C} R_n^{-1}$$

$=$  $=: H$

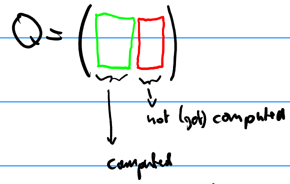
Fact:

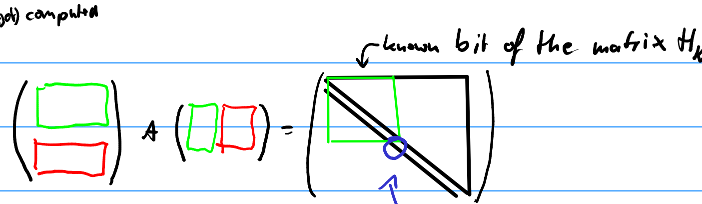


Have: $A Q_n = Q_n H \quad \overset{\Rightarrow}{\text{orth.}} \quad q_i^T A q_j = H_{ij}$

→ Arnoldi iteration   DEMO pt. 1

Eigenvalues?



$H = Q^T A Q =$ 

Eigenvalues of H_k vs eigenvalues of H?

eigenvalues of H_k ("Ritz values")

$\longrightarrow$ eigenvalues of H_n = eigenvalues of A

[Eigenvalues of H_k still need to be computed. How?

DEMO pt. 2

[Somewhat expensive – less expensive for symmetric

→ Lanczos iteration   (see hw)