Welcome to Math 221! Lecture 19, Fall 24

Symmetric Eigenproblem + SVD

Real: $A = A^T = Q \Lambda Q^T$, $QQ^T = I$

$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$

$\lambda_1 \geq \cdots \geq \lambda_n$

$Q = [q_1, \dots, q_n]$

Def: Rayleigh Quotient $\rho(v, A) = \frac{v^T A v}{v^T v}$, $v \neq 0$

Props: $Av = \lambda v \Rightarrow \rho(v, A) = \lambda$

$v = QQ^T v = Qb \Rightarrow \rho(v, A) = \frac{b^T \Lambda b}{b^T b}$

= convex comb. of evals

Courant-Fischer Minimax Thm

R^j = j-dim subspace of $\mathbb{R}^n$

S^{n-j+1} = $n-j+1$-dim subspace of $\mathbb{R}^n$

$$\max_{R^j} \min_{\substack{r \in R^j \\ r \neq 0}} \rho(r, A) = \lambda_j = \min_{S^{n-j+1}} \max_{\substack{s \in S^{n-j+1} \\ s \neq 0}} \rho(s, A)$$

Weyl's Thm: $A = A^T$ with $\lambda_1, \dots, \lambda_n$ and $E = E^T$

$A + E$ has evals $\nu_1 \geq \cdots \geq \nu_n$ with

$|\nu_i - \lambda_i| \leq \|E\|_2 \quad \forall i$

Corollary for SVD: if A general,

sing vals $\sigma_1 \geq \cdots \geq \sigma_n$

and $A+E$ has sing vals $\tau_1 \geq \cdots \geq \tau_n$

then $|\sigma_i - \tau_i| \leq \|E\|_2$

[follows from $\begin{bmatrix} 0 & A+E \\ A^T+E^T & 0 \end{bmatrix}$]

proof of Weyl:

$$\mu_j = \min_{S^{n-j+1}} \max_{\substack{s \in S^{n-j+1} \\ s \neq 0}} \frac{s^T(A+E)s}{s^Ts}$$

$$= \min \max \frac{s^TAs}{s^Ts} + \frac{s^TEs}{s^Ts}$$

$$\leq \min \max \frac{s^TAs}{s^Ts} + \|E\|_2$$

$$= \lambda_j + \|E\|_2$$

for other inequality, swap λ_j and μ_j

Def: Inertia(A) =

$(\#\text{neg\_evals}(A), \#\text{zero\_evals}(A), \#\text{pos\_evals}(A))$

Sylvester's Thm: $A = A^T$, X nonsingular

$\Rightarrow$ Inertia(A) = Inertia(X^TAX)

Fact: Suppose we factor $A = LDL^T$

(Gauss elim with symmetric or no pivoting)

Inertia (A) = Inertia (D)

$= (\#D_{ii} < 0, \#D_{ii} = 0, \#D_{ii} > 0)$

Factor $A - xI = L' \cdot D' \cdot L'^T$

$\#D'_{ii} < 0 = \#$ evals of $A - xI < 0$

$= \#$ evals of $A < x$

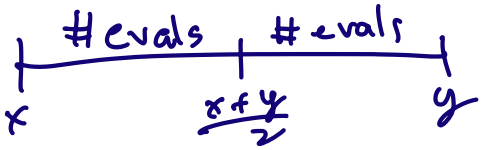
Factor $A - yI$, $x < y$

get $\#$ evals of $A \leq y$

$\#$ evals in $[x, y] = (\#\text{evals} \leq y) - (\#\text{evals} < x)$

$\Rightarrow$ count $\#$ evals in any interval

$\Rightarrow$ bisection works: Inertia $\left(A - \frac{x+y}{2} I\right)$



in practice, start by reducing $A = QTQ^T$

to tridiagonal T, then

computing inertia $(T - xI)$ costs $O(n)$

Proof of Sylvester's Thm

Suppose $\#$ evals of $A < 0$ is m

but $\#$ evals of $X^T A X < 0$ is $m' < m$

seek contradiction

N = m-dimensional subspace spanned by evecs of m negative evals of A

$\Rightarrow 0 \neq x \in N \Rightarrow x^T A x < 0$

P = $n-m'$ dimensional subspace spanned by $n-m'$ evecs of $n-m'$ nonnegative evals of $X^T A X$

$\Rightarrow 0 \neq x \in P \Rightarrow x^T X^T A X x = (Xx)^T A (Xx) \geq 0$

($Xx \in XP$)

$\dim(XP) + \dim(N)$

$= n - m' + m > n$

$\Rightarrow$ they intersect in nonzero z

$\Rightarrow z^T A z \geq 0$ and $z^T A z < 0$

contradiction

Perturbation Theory for Evecs

Thm: $A = Q \Lambda Q^T$ $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$

$A + E = Q' \Lambda' Q'^T$ $\Lambda' = \mathrm{diag}(\lambda_1', \ldots, \lambda_n')$

$\theta_i = \mathrm{angle}(q_i, q_i')$

$\mathrm{gap}(i, A) = \min_{j \neq i} |\lambda_i - \lambda_j|$

$$|\sin(2\theta_i)| \leq \frac{\|E\|_2}{\mathrm{gap}(i, A)}$$

$\sim \theta_i$ if $\theta_i \ll 1$

Why gap in denominator?

if $A=I$, gap $=0$, worst case:

upper bound $=\infty$

$$\begin{bmatrix} 1.1 & 0 \\ 0 & .9 \end{bmatrix} \quad \text{gap} = .2$$

$$E = \begin{bmatrix} -.1 & \\ & +.1 \end{bmatrix} \Rightarrow A+E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\frac{\|E\|_2}{\text{gap}} = \frac{.1}{.2} = \frac{1}{2}$$

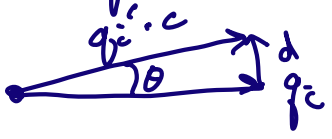
$$|.5\sin(2\theta)| \le \frac{1}{2}$$

$|\sin(2\theta)| \le 1$ no control on θ

proof of weaker result (full proof in text)

write evec of $A+E$ as $q_i + d$

where $d^T q_i = 0$



$q_i' = \frac{q_i + d}{\|q_i + d\|_2}$: goal: bound $\|d\|_2 = \tan\theta$

$$(A+E)(q_i+d) = \lambda_i'(q_i+d)$$

$$Aq_i + Ad + Eq_i + \underline{Ed} = \lambda_i' q_i + \lambda_i' d$$

ignore second order term

$$LHS = (A + E - \lambda_i' I) q_i = (\lambda_i' I - A) d = RHS$$

$$d = \sum_{j \neq i} d_j q_j \qquad \|d\|_2 = \sqrt{\sum_{j \neq i} d_j^2}$$

$$LHS = (A + E - \lambda_i' I) q_i$$

$$= (\lambda_i I + E - \lambda_i' I) q_i$$

$$= ((\lambda_i - \lambda_i') I + E) q_i$$

$$\|LHS\|_2 \leq \|(\lambda_i - \lambda_i') q_i\|_2 + \|E q_i\|_2 \qquad \|q_i\|_2 = 1$$

$$\leq \underset{\text{Weyl}}{\|E\|_2} + \|E\|_2 = 2\|E\|_2$$

$$RHS = (\lambda_i' I - A) \sum_{j \neq i} d_j q_j$$

$$= \sum_{j \neq i} (\lambda_i' d_j q_j - \lambda_j d_j q_j)$$

$$\lambda_i' - \lambda_j = \underbrace{\lambda_i' - \lambda_i}_{\leq \|E\|_2 \text{ Weyl}} + \underbrace{\lambda_i - \lambda_j}_{\geq gap}$$

$$\sum_{j \neq i} (\lambda_i' - \lambda_j) d_j q_j .$$

$$\|RHS\|_2 \geq (gap - \|E\|) \|d\|_2$$

$$\frac{2\|E\|}{gap} \approx \frac{2\|E\|}{gap - \|E\|} \geq \|d\|_2 = \tan\theta$$

if $\|E\| < \frac{1}{2} gap$ QED

More results on Rayleigh Quotient: why good eval approximation:

Thm: Given $\|x\|_2 = 1$ and β

Then A has eval α:

$$|\alpha - \beta| \leq \|Ax - \beta x\|_2$$

Given only x, $\beta = \rho(x, A)$

minimizes $\|Ax - \beta x\|_2$

Given any unit vector x, $\exists$ eval within distance $\|Ax - \rho(x,A)x\|_2$ of $\rho(x,A)$ and $\rho(x,A)$ minimizes distance

Now let λ_i be eval of A closest to $\rho(x,A)$ and $gap = \min_{j \neq i} |\lambda_j - \rho(x,A)|$

then $|\lambda_i - \rho(x,A)| \leq \dfrac{\|Ax - \rho(x,A)x\|_2^2}{gap}$

(result later: cubic convergence in shifted QR)

proof: Part 1: $\|x\|_2 = 1$

$$1 = \|x\|_2 = \|(A - \beta I)^{-1}(A - \beta I)x\|_2$$

$$\leq \|(A - \beta I)^{-1}\|_2 \cdot \|(A - \beta I)x\|_2$$

$A = Q \Lambda Q^T$

$$= \| (\Lambda - \beta I)^{-1} \|_2 \cdot \| Ax - \beta x \|$$

$$= \frac{1}{\min_i |\lambda_i - \beta|} \cdot \| Ax - \beta x \|$$

$$\min_i |\lambda_i - \beta| \leq \| Ax - \beta x \|_2$$

Part 2: show that $\beta = \rho(x, A)$ minimizes $\|Ax - \beta x\|_2$

$$Ax - \beta x = \underbrace{Ax - \rho(x,A)x}_{y} + \underbrace{\rho(x,A)x - \beta x}_{z}$$

if $z^T y = 0 \Rightarrow$

$$\|Ax - \beta x\|_2^2 = \|y\|_2^2 + \|z\|_2^2$$

$$\geq \|y\|_2^2 = \|Ax - \rho(x,A)x\|_2^2$$

$$z^T y = (\rho(x,A) - \beta) x^T (Ax - \rho(x,A)x)$$

$$= (\rho(x,A) - \beta)(x^T Ax - \rho(x,A) x^T x)$$

$$= 0$$

by def of $\rho(x, A)$

Part 3: consider special case of 2x2 diagonal (captures all ideas of full proof):

$$A = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \quad x = \begin{bmatrix} c \\ s \end{bmatrix} \quad c^2 + s^2 = 1$$

$$\rho(x, A) = c^2 \lambda_1 + s^2 \lambda_2$$

assume $c > s \Rightarrow \rho(x, A)$ closer to λ_1

can show $$\frac{\|Ax - \rho(x,A)x\|_2^2}{gap} = |\lambda_1 - \rho(x, A)|$$

vs $\leq$ in general case

Algorithms: Design Space

(1) "Usual Accuracy": backward stable

exact evals, evecs of $A+E$, $\|E\|_2 = O(\varepsilon) \cdot \|A\|_2$

(1.1) all evals (with or without evecs)

(1.2) just evals in $[x, y]$ (w or w/o evecs)

(1.3) just evals $\lambda_i, \lambda_{i+1}, \ldots \lambda_j$

eg $\lambda_1, \ldots \lambda_{10}$: 10 largest evals (w or w/o evecs)

(1.2) and (1.3) can be much cheaper than (1.1) when only few evals/evecs desired

(2) "High accuracy": get tiny evals (and their evecs) with more leading digits

Ex: if A well conditioned

i.e all evals ~ same magnitude

than usual accuracy → error bound is $O(\varepsilon) \cdot \|A\|$ ⇒ all computed evals have correct leading digits

What structures let us compute tiny sing values or evals accurately?

$B = D \cdot A$, $D = \text{diag}(d_1, \ldots, d_n)$

some $d_i >>$ other d_j

⇒ some $\sigma_i >>$ other σ_j

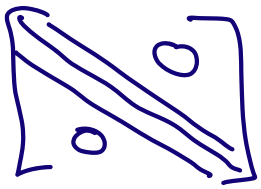
⇒ usual accuracy guarantees no correct digits for small σ_j

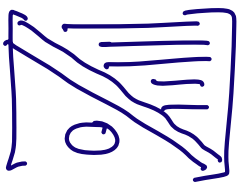
∃ perturbation theory and algs that say small normwise perturbation to A keeps all σ_i to high relative accuray if A well conditioned

Symmetric case: DAD same idea

Algorithms and costs for case (1)

(1) Reduce $A = QTQ^T$ $QQ^T = I$

$T = T^T$ tridiagonal 

same idea as Hessenberg reduction in Chap4 $A = QHQ^T =$ 

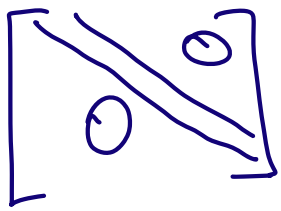
(routine in LAPACK: ssytrd)

There is an alg for tridiagal reduction that moves $O\left(\frac{n^3}{\sqrt{M}}\right)$ words

If A banded. 

cost drops from $O(n^3)$ to $O(n^2 \cdot bw)$

- SVD analogous: $B = UAV^T$
  - B: bidiagonal
  - U, V: orthogonal

$B =$ 

(1.1) Given T: find all evals,
w or w/o evecs
many algs, all cost $O(n^2)$ for
evals alone, If evecs desired
costs range from $O(n^2)$ to $O(n^3)$,
varying numerical stability

(1.1.1) Oldest is QR iteration
with shift as in Chap 4

Thm (Wilkinson) with right shift,
tridiagonal QR is globally convergent,
usually cubic (# digits triples
at every iteration)

Cost: $O(n^2)$ for evals alone,
but $O(n^3)$ for evecs too
Lapack: ssyev (sstegr)

Variant for SVD (Golub + Kahan,
1965)

LAPACK uses a variant (for
biadiagonal SVD) guaranteeing
high relative accuracy in all
sing values (D., Kahan)

(1.1.2) Improve cost of $O(n^3)$ for evecs to $O(n^2)$, but not guarantee orthogonal evecs if evals close:

$\frac{\|Ax - \lambda x\|}{\|A\|}$ will be $O(\varepsilon)$, but $x_i^T x_j$ may be large

(1) compute evals: $O(n^2)$... LAPACK: sstebz

(2) compute each evec using inverse iteration (Chap 4)

$$x_{i+1} = (T - \lambda I)^{-1} x_i$$

LAPACK: sstein

fast convergence if λ accurate

if λ, λ' very close, no guarantee of orthogonality; worst case if two evals round to same floating point number

Long term Goal: $O(n^2)$ and orthogonality: solved by MRRR (Parlett) LAPACK: ssyevr (sstemr)

(1.1.3) Divide + conquer:
(Cuppen, Gu+Eisentat)

faster than QR, not as

fast as MRRR or inverse iteration, guarantees orthogonal evecs: cost $O(n^g)$, $2 < g < 3$

same idea is used for updating evals evecs of $A + xx^T$

LAPACK: ssyevd (sstedc)