Welcome to Ma221! (Mar 29)

Symmetric Eigenproblem + SVD

Goals: Perturbation Theory
Algorithms

Perturbation theory: needed to understand algorithms

Real symmetric $A = A^T$

Complex Hermitian $A = A^H$

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

$$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n) \quad \lambda_1 \geq \cdots \geq \lambda_n$$

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

Complex symmetric different

$$\begin{bmatrix} 1 & i \\ i & -1 \end{bmatrix} \quad i = \sqrt{-1}$$

double eigenvalue at 0
not diagonalizable

Most results will apply to SVD
(Thm 3.3 part 4)

$$B = \left[\begin{array}{c|c} 0 & A \\ \hline A^T & 0 \end{array}\right] = B^T$$

eigendecomp of B related to svd of A

evals of B = $\pm$ sing vals of A
(+ zeros if A rectangular)

evecs closely related to sing vecs of A

reuse algorithms for eig(B) to get
SVD(A)
(some open problems)

Small perturbation of A to $A+E$
changes B to $B + \left[\begin{array}{c|c} 0 & E \\ \hline E^T & 0 \end{array}\right]$

$\Rightarrow$ use perturbation theory for sym eig
for SVD

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

Properties: if $Au = \lambda u \Rightarrow \rho(u, A) = \lambda$

$$u = \sum_{i=1}^{n} b_i q_i = Qb \,, \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

$$\rho(u, A) = \frac{(Qb)^T A (Qb)}{(Qb)^T (Qb)} = \frac{b^T \overbrace{Q^T A Q}^{\Lambda} b}{b^T \underbrace{Q^T Q}_{=I} b}$$

$$= \frac{b^T \Lambda b}{b^T b}$$

$$= \sum_{i=1}^{n} \lambda_i b_i^2 \Big/ \sum_{i=1}^{n} b_i^2$$

= convex combination of all evals

$$\lambda_1 \ge \rho(v, A) \ge \lambda_n$$

$$\lambda_1 = \max_{v \ne 0} \rho(v, A), \quad \lambda_n = \min_{v \ne 0} \rho(v, A)$$

All evals expressed using $\rho(v, A)$

Courant-Fischer Minimax Thm

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

S^{n-j+1} = $n-j+1$ " " " "

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

max over R^j attained by $\text{span}(q_1, \ldots, q_j)$

min over $r \in R^j$ " " $r = q_j$

min over S^{n-j+1} " " $\text{span}(q_j, q_{j+1} \cdots q_n)$

max over $s \in S^{n-j+1}$ " " $s = q_j$

Proof: Given R^j and S^{n-j+1}

their dimensions add up to $j + n - j + 1 = n + 1$

$\Rightarrow$ must intersect in some nonzero x_{RS}

$$\min_{\substack{r \in R^j \\ r \ne 0}} \rho(r, A) \le \rho(x_{RS}, A) \le \max_{\substack{s \in S^{n-j+1} \\ s \ne 0}} \rho(s, A)$$

Let R' maximize $\min_{\substack{r \in R' \\ r \neq 0}} \rho(r, A)$, $\dim R' = j$

Let S' minimize $\max_{s \in S'} \rho(s, A)$, $\dim S' = n-j+1$

(*) $\lambda_j \le \max_{R^j} \min_{\substack{r \in R' \\ r \neq 0}} \rho(r, A) = \min_{\substack{r \in R' \\ r \neq 0}} \rho(r, A) \le \rho(x_{R'S'})$

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

If we choose $R^j = \mathrm{span}(q_1, \dots, q_j)$

$r = q_j \Rightarrow \min_{r \in R^j} \rho(r, A) = \rho(q_j, A) = \lambda_j$

If we choose $S^{n-j+1} = \mathrm{span}(q_j \cdots q_n)$

and $s = q_j$ then

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

$\Rightarrow$ all inequalities in (*) are equalities

Weyl's Thm $A = A^T$ with eval $\lambda_1 \ge \dots \ge \lambda_n$

and $E = E^T$ where

$A + E$ has evals $\mu_1 \ge \dots \ge \mu_n$ then

$|\lambda_i - \mu_i| \le \|E\|_2$ for all i

Corollary for SVD: if A general with sing vals $\sigma_1 \geq \cdots \geq \sigma_n$ and $A+E$ has sing vals $\tau_1 \geq \cdots \geq \tau_n$ then $|\sigma_i - \tau_i| \leq \|E\|_2$ for all i

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

Proof of Weyl:

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

$$= \min \max \left(\frac{s^T \overset{A+E}{A} s}{s^T s} + \underbrace{\frac{s^T \overset{-E}{E} s}{s^T s}}_{\leq \|E\|_2} \right)$$

$\mu_j \quad + \|E\|_2$

$$\leq \underbrace{\min \max \frac{s^T A s}{s^T s}}_{=\ \lambda_j} + \|E\|_2$$

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

swap roles of μ_j and λ_j to get other inequality

$$\lambda_j \leq \mu_j + \|E\|_2 \Rightarrow |\lambda_j - \mu_j| \leq \|E\|_2$$

Def: Inertia(A) =
(# neg evals (A),
zero evals (A),
pos evals (A))

Sylvester's Thm: $A = A^T$, X nonsingular
$\Rightarrow$ Inertia(A) = Inertia$(X^T A X)$

Fact: Suppose we factor $A = LDL^T$
(Gaussian elim with symmetric or no pivoting)
Inertia(A) = Inertia(D)
$= (\# D_{ii} < 0, \# D_{ii} = 0, \# D_{ii} > 0)$

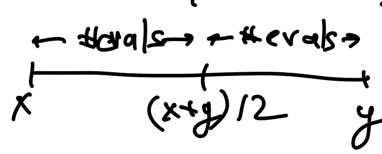
Factor $A - xI = L' \cdot D' \cdot L'^T$
$\# D'_{ii} < 0 \Rightarrow$ # evals of $A - xI < 0$
= # evals of $A < x$

Factor $A - yI$, compute $(x < y)$
evals of $A < y$

evals of A in $[x, y]$ =
evals $\leq y$ − # evals $< x$

$\Rightarrow$ count # evals in any interval

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



keep bisecting interval until
as narrow as desired
more details on implementation of
"bisection" later: starts by
reducing $A = QTQ^T$, T tridiagonal
bisection on T, cost $= O(n)$

Proof of Sylvester's Thm

Suppose # evals of $A < 0$ is m
and # evals of $X^T A X < 0$ is m' but $m' < m$
seek contradiction

N = m-dimensional subspace for
m negative evals of A

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

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

$\Rightarrow x \in P \Rightarrow x^T X^T A \underbrace{(Xx)}_{=y} \geq 0$

$y \in XP \Rightarrow y^T A y \geq 0$

$\dim(XP) + \dim(N) =$
$n - m' + m > n$

$\Rightarrow$ X^P and N intersect in $z \neq 0$

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

contradiction

Perturbation Theory for Evecs

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

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

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

$Q' = [q_1', \ldots, q_n']$

θ_i = angle between q_i and q_i'

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

Then: $$\left| .5 \sin(2\theta_i) \right| \leq \frac{\|E\|_2}{gap(i,A)}$$

$\sim \theta_i$

if $\theta_i << 0$

Why $gap(i,A)$ in denominator

Worst case $A = I$: $gap(i,A) = 0$

upper bound $= \infty$

Proof of a weaker result

(see text for full result)

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

where $d^T q_i = 0$

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

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

$$Aq_i + Ad + Eq_i + \underbrace{Ed}_{\text{ignore second order term}} = \lambda_i' q_i + \lambda_i' d$$

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

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

$$d = \sum_{j \neq i} d_j q_j$$

$$2\|E\|_2 \geq LHS = (\lambda_i I + E - \lambda_i' I)q_i = \sum_{j\neq i} (\lambda_i' - \lambda_j) d_j q_j = RHS \geq (gap - \|E\|) \cdot \|d\|$$

$\|LHS\|_2 \leq 2\|E\|_2$ because $|\lambda_i - \lambda_i'| \leq \|E\|_2$ by Weyl

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

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

$$\frac{2\|E\|_2}{gap} \approx \frac{2\|E\|_2}{gap - \|E\|_2} \geq \|d\|_2 = |\tan\theta| \sim |\theta| \text{ if } |\theta| \ll 1$$

if $\|E\|_2 \ll gap$

More results on why Rayleigh Quotient good approx eval

$$\rho(u,A) = \frac{u^T A u}{u^T u} \qquad u \neq 0$$

Thm Given $\|x\|_2 = 1$ and β

Then A has an eval α

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

Given only x, $\beta = \rho(x,A)$ minimizes $\|Ax - \beta x\|_2$

Given any unit vector x, there is an eval within distance $\|Ax - \rho(x,A) \cdot x\|_2$ of $\rho(x,A)$ and $\rho(x,A)$ minimizes this distance

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

Then $|\lambda_i - \rho(x,A)| \le \dfrac{\|Ax - \rho(x,A)x\|_2^2}{\text{gap}}$

(later: get cubic convergence in QR iteration)

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

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

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

write $A = Q\Lambda Q^T$

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

$$= \frac{1}{\min_i |\lambda_i - \beta|} \cdot \|Ax - \beta x\|_2$$

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

Part 2: to 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 $y^T z = 0$

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

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

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

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

$= 0$ by def of $\rho(x,A)$

Last Part: Do special case of 2x2 diagonal matrix, captures all ideas of general case

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

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

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

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

$\leq$ in general case