Symmetric Eigenvalue Problems

- The symmetric eigenvalue problem: basic facts
- Min-Max theorem -
- Inertia of matrices
- Bisection algorithm
- QR algorithm for symmetric matrices
- The Jacobi method

The symmetric eigenvalue problem: Basic facts

➤ Consider the Schur form of a real symmetric matrix A:

$$A = QRQ^H$$

Since $A^H = A$ then $R = R^H$ ➤

Eigenvalues of A are real

and

There is an orthonormal basis of eigenvectors of A

In addition, Q can be taken to be real when A is real.

$$(A - \lambda I)(u + iv) = 0 \rightarrow (A - \lambda I)u = 0 \;\&\; (A - \lambda I)v = 0$$

➤ Can select eigenvector to be either u or v

The min-max theorem (Courant-Fischer)

Label eigenvalues decreasingly: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$

The eigenvalues of a Hermitian matrix A are characterized by the relation

$$\lambda_k = \max_{S,\ \dim(S)=k} \quad \min_{x \in S, x \neq 0} \frac{(Ax, x)}{(x, x)}$$

Proof: Preparation: Since A is symmetric real (or Hermitian complex) there is an orthonormal basis of eigenvectors $u_1, u_2, \cdots, u_n$. Express any vector x in this basis as $x = \sum_{i=1}^n \alpha_i u_i$. Then : $(Ax, x)/(x, x) = [\sum \lambda_i |\alpha_i|^2]/[\sum |\alpha_i|^2]$.

(a) Let S be any subspace of dimension k and let $\mathcal{W} = \text{span}\{u_k, u_{k+1}, \cdots, u_n\}$. A dimension argument (used before) shows that $S \cap \mathcal{W} \neq \{0\}$. So there is a non-zero x_w in $S \cap \mathcal{W}$.

➤ Express this x_w in the eigenbasis as $x_w = \sum_{i=k}^{n} \alpha_i u_i$. Then since $\lambda_i \le \lambda_k$ for $i \ge k$ we have:

$$\frac{(Ax_w, x_w)}{(x_w, x_w)} = \frac{\sum_{i=k}^{n} \lambda_i |\alpha_i|^2}{\sum_{i=k}^{n} |\alpha_i|^2} \le \lambda_k$$

Thus, for any subspace S of dim. k we have $\min_{x\in S, x\neq 0}(Ax, x)/(x, x) \le \lambda_k$.

(b) We now take $S_* = \mathbf{span}\{u_1, u_2, \cdots, u_k\}$. Since $\lambda_i \ge \lambda_k$ for $i \le k$, for this particular subspace we have:

$$\min_{x\,\in\,S_*,\;x\neq 0} \frac{(Ax, x)}{(x, x)} = \min_{x\,\in\,S_*,\;x\neq 0} \frac{\sum_{i=1}^{k} \lambda_i |\alpha_i|^2}{\sum_{i=k}^{n} |\alpha_i|^2} = \lambda_k.$$

(c) The results of (a) and (b) imply that the max over all subspaces S of dim. k of $\min_{x\in S, x\neq 0}(Ax, x)/(x, x)$ is equal to λ_k □

- Consequences:

$$\lambda_1 = \max_{x \neq 0} \frac{(Ax, x)}{(x, x)} \qquad \lambda_n = \min_{x \neq 0} \frac{(Ax, x)}{(x, x)}$$

- Actually 4 versions of the same theorem. 2nd version:

$$\lambda_k = \min_{S,\ \dim(S)=n-k+1} \quad \max_{x \in S, x \neq 0} \frac{(Ax, x)}{(x, x)}$$

- Other 2 versions come from ordering eigenvalues increasingly instead of decreasingly.

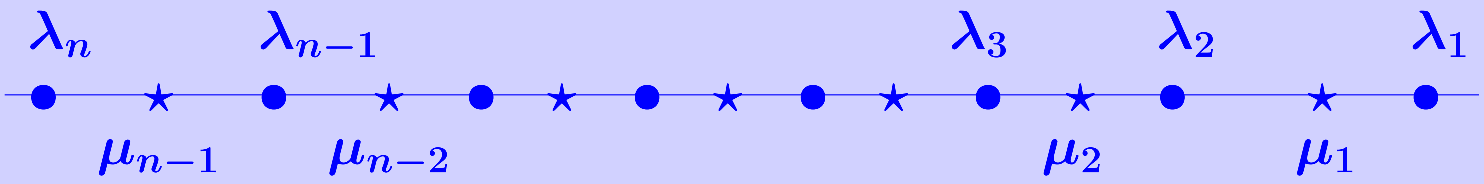
✍1 Write down all 4 versions of the theorem

✍2 Use the min-max theorem to show that $\|A\|_2 = \sigma_1(A)$ - the largest singular value of A.

➤ Interlacing Theorem: Denote the $k \times k$ principal submatrix of A as A_k, with eigenvalues $\{\lambda_i^{[k]}\}_{i=1}^k$. Then

$$\lambda_1^{[k]} \geq \lambda_1^{[k-1]} \geq \lambda_2^{[k]} \geq \lambda_2^{[k-1]} \geq \cdots \lambda_{k-1}^{[k-1]} \geq \lambda_k^{[k]}$$

Example: λ_i's = eigenvalues of A, μ_i's = eigenvalues of A_{n-1}:



➤ Many uses.

➤ For example: interlacing theorem for roots of orthogonal polynomials

The Law of inertia (real symmetric matrices)

➤ Inertia of a matrix = [m, z, p] with $m =$ number of < 0 eigenvalues, $z =$ number of zero eigenvalues, and $p =$ number of > 0 eigenvalues.

Sylvester's Law of inertia:

If $X \in \mathbb{R}^{n\times n}$ is nonsingular, then A and X^TAX have the same inertia.

➤ Terminology: X^TAX is congruent to A

✍3 Suppose that $A = LDL^T$ where L is unit lower triangular, and D diagonal. How many negative eigenvalues does A have?

✍4 Assume that A is tridiagonal. How many operations are required to determine the number of negative eigenvalues of A?

✍5 Devise an algorithm based on the inertia theorem to compute the i-th eigenvalue of a tridiagonal matrix.

✍6 Let $F \in \mathbb{R}^{m\times n}$, with $n < m$, and F of rank n.

What is the inertia of the matrix on the right:
[Hint: use a block LU factorization]

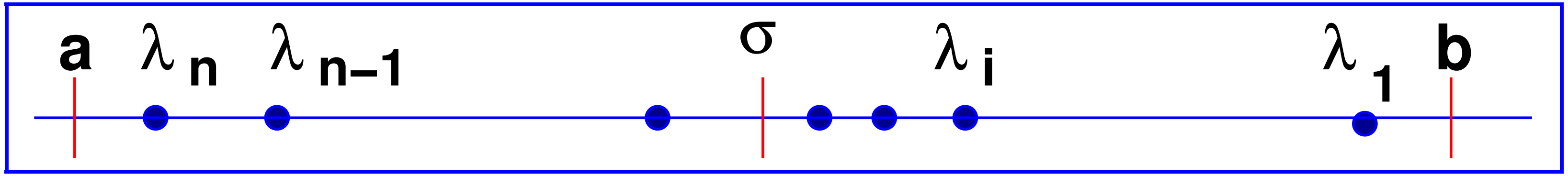
$$\begin{pmatrix} I & F \\ F^T & 0 \end{pmatrix}$$

➤ Note 1: Converse result also true: If A and B have same inertia they are congruent. [This part is easy to show]

➤ Note 2: result also true for (complex) Hermitian matrices ($X^H A X$ has same inertia as A).

Bisection algorithm for tridiagonal matrices:

- Goal: to compute i-th eigenvalue of A (tridiagonal)
- Get interval $[a, b]$ containing spectrum [Gerschgorin]: $a \le \lambda_n \le \cdots \le \lambda_1 \le b$
- Let $\sigma = (a + b)/2$ = middle of interval
- Calculate $p =$ number of positive eigenvalues of $A - \sigma I$
  - If $p \ge i$ then $\lambda_i \in \ (\sigma, \ b] \to$ set $a := \sigma$



  - Else then $\lambda_i \in \ [a, \ \sigma] \to$ set $b := \sigma$
- Repeat until $b - a$ is small enough.

The QR algorithm for symmetric matrices

➤ Most important method used : reduce to tridiagonal form and apply the QR algorithm with shifts.

➤ Householder transformation to Hessenberg form yields a tridiagonal matrix because

$$HAH^T = A_1$$

is symmetric and also of Hessenberg form ➤ it is tridiagonal symmetric.

Tridiagonal form preserved by QR similarity transformation

Practical method

➤ How to implement the QR algorithm with shifts?

➤ It is best to use Givens rotations – can do a shifted QR step without explicitly shifting the matrix..

➤ Two most popular shifts:

$$s = a_{nn} \text{ and } s = \text{smallest e.v. of } A(n-1:n, n-1:n)$$

Jacobi iteration - Symmetric matrices

➤ Main idea: Rotation matrices of the form

$$J(p,q,\theta) = \begin{pmatrix} 1 & \dots & 0 & & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & c & \cdots & s & \cdots & 0 \\ \vdots & \cdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & -s & \cdots & c & \cdots & 0 \\ \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & \dots & 0 & & \dots & & 1 \end{pmatrix} \begin{matrix} \\ \\ p \\ \\ q \\ \\ \\ \end{matrix}$$

$c = \cos\theta$ and $s = \sin\theta$ are so that $J(p,q,\theta)^T A J(p,q,\theta)$ has a zero in position (p,q) (and also (q,p))

➤ Frobenius norm of matrix is preserved – but diagonal elements become larger ➤ convergence to a diagonal.

- Let $B = J^T A J$ (where $J \equiv J_{p,q,\theta}$).
- Look at 2×2 matrix $B([p,q],[p,q])$ (matlab notation)
- Keep in mind that $a_{pq} = a_{qp}$ and $b_{pq} = b_{qp}$

$$\begin{pmatrix} b_{pp} & b_{pq} \\ b_{qp} & b_{qq} \end{pmatrix} = \begin{pmatrix} c & -s \\ s & c \end{pmatrix} \begin{pmatrix} a_{pp} & a_{pq} \\ a_{qp} & a_{qq} \end{pmatrix} \begin{pmatrix} c & s \\ -s & c \end{pmatrix} = \dots$$

$$= \left[\begin{array}{c|c} c^2 a_{pp} + s^2 a_{qq} - 2sc\, a_{pq} & (c^2 - s^2) a_{pq} - sc(a_{qq} - a_{pp}) \\ \hline * & c^2 a_{qq} + s^2 a_{pp} + 2sc\, a_{pq} \end{array} \right]$$

- Want:

$$(c^2 - s^2) a_{pq} - sc(a_{qq} - a_{pp}) = 0$$

$$\frac{c^2 - s^2}{2sc} = \frac{a_{qq} - a_{pp}}{2a_{pq}} \equiv \tau$$

- Letting $t = s/c \ (= \tan\theta)$ $\quad \rightarrow$ quad. equation

$$t^2 + 2\tau t - 1 = 0$$

- $t = -\tau \pm \sqrt{1+\tau^2} = \frac{1}{\tau \pm \sqrt{1+\tau^2}}$
- Select sign to get a smaller t so $\theta \le \pi/4$.
- Then : $\quad c = \frac{1}{\sqrt{1+t^2}}; \qquad s = c * t$
- Implemented in matlab script `jacrot(A,p,q)` –

➤ Define: $A_O = A - \text{Diag}(A)$ $\equiv$ A 'with its diagonal entries replaced by zeros'

➤ Observations: (1) Unitary transformations preserve $\|.\|_F$. (2) Only changes are in rows and columns p and q.

➤ Let $B = J^T AJ$ (where $J \equiv J_{p,q,\theta}$). Then:

$$a_{pp}^2 + a_{qq}^2 + 2a_{pq}^2 = b_{pp}^2 + b_{qq}^2 + 2b_{pq}^2 = b_{pp}^2 + b_{qq}^2$$

because $b_{pq} = 0$. Then, a little calculation leads to:

$$\begin{aligned}
\|B_O\|_F^2 &= \|B\|_F^2 - \sum b_{ii}^2 = \|A\|_F^2 - \sum b_{ii}^2 \\
&= \|A\|_F^2 - \sum a_{ii}^2 + \sum a_{ii}^2 - \sum b_{ii}^2 \\
&= \|A_O\|_F^2 + (a_{pp}^2 + a_{qq}^2 - b_{pp}^2 - b_{qq}^2) \\
&= \|A_O\|_F^2 - 2a_{pq}^2
\end{aligned}$$

➤ $\|A_O\|_F$ will decrease from one step to the next.

✍7 Let $\|A_O\|_I = \max_{i \neq j} |a_{ij}|$. Show that

$$\|A_O\|_F \leq \sqrt{n(n-1)}\|A_O\|_I$$

✍8 Use this to show convergence in the case when largest entry is zeroed at each step.