MAI0119/Lecture 5.5 - Contents

Perturbation Theory

Theorem *(Bauer-Fike)* If μ is an eigenvalue of $A + E \in \mathbb{C}^{n\times n}$, A is non-defective, then

$$\min_{\lambda\in\lambda(A)} |\lambda - \mu| \leq \kappa_p(X)\|E\|_p,$$

where $\|\cdot\|_p$ denotes any of the p-norms and X is the eigenvector matrix of A.

Remark The QR algorithm computes a Schur decomposition

$$\hat{T} = Q^T(A+E)Q, \quad Q^TQ = I, \quad \|E\|_2 \leq \mathcal{O}(n\mathbf{u})\|A\|_2.$$

The largest eigenvalues are computed with good relative accuracy.

Corollary Let $A \in \mathbb{C}^{n\times n}$ be non–defective with eigenvector matrix X. There is an eigenvalue λ of A such that

$$|\lambda - \hat{\lambda}| \leq \kappa_2(X)\|r\|_2.$$

Proof Let $(\hat{x}, \hat{\lambda})$ be an *approximate* eigenpair of A, with $\|\hat{x}\|_2 = 1$, and put $E = r\hat{x}^H$, where $r = A\hat{x} - \hat{\lambda}\hat{x}$.

Conditioning for a single eigenvalue

Definition Let λ be an eigenvalue of $A \in \mathbb{C}^{n\times n}$. If $y^HA = \lambda y^Y$, $\|y\|_2 = 1$, then y is a *left eigenvector* of A.

Remark The existance of left and right eigenvectors follows from the Jordan decomposition.

Lemma Suppose λ is a simple eigenvalue of $A \in \mathbb{C}^{n\times n}$. The left- and right eigenvectors satisfy $y^Hx \neq 0$.

Lemma Let (λ, x) be a simple eigenvalue of $A \in \mathbb{C}^{n\times n}$ and $A(t) = A + tE$. Then

$$\lambda(t) = \lambda + ty^H Ex + \mathcal{O}(t^2),$$

where y is the left-eigenvector associated with λ.

Definition The condition number for a simple eigenvalue λ is

$$\kappa_2(\lambda, A) = \frac{\|x\|_2\|y\|_2}{|y^H x|}.$$

The Symmetric Eigenvalue Problem

Theorem If $A \in \mathbb{R}^{n\times n}$ is *symmetric* then there exists a real orthogonal Q such that

$$Q^T AQ = D = \mathrm{diag}(\lambda_1, \dots, \lambda_n).$$

Remark This follows directly from the *Schur decomposition*.

The eigenvalues are real since

$$\lambda x^H x = x^H(\lambda x) = x^H(Ax) = (A^H x)^H x = (Ax)^H x = (\lambda x)^H x = \bar{\lambda} x^H x$$

Theorem If $A \in \mathbb{R}^{n\times n}$ is symmetric then

$$\lambda_k(A) = \max_{\dim(S)=k} \min_{y\in S, \neq 0} \frac{y^T Ay}{y^T y}, \qquad k = 1, 2, \dots, n. \quad \square$$

Remarks This is called the *Courant–Fischer Minimax theorem*.

The Law of Inertia

Definition The *Inertia* of a symmetric matrix A is a triplet (m, z, p) where m, z, and p, are the number of positive, zero, and negative eigenvalues respectively.

Theorem *(Sylvester's Law)* If the matrix A is symmetric and X is non-singular then A and $X^T AX$ have the same inertia.

Remark Subtract a shift and compute the decomposition

$$A - \mu I = LDL^T,$$

to find out how many eigenvalues λ_i are larger or smaller than μ.

Tridiagonal Methods

Let,

$$T = \begin{pmatrix} a_1 & b_1 & 0 & 0 & 0 \\ b_1 & a_2 & b_2 & 0 & 0 \\ 0 & b_2 & a_3 & b_3 & 0 \\ 0 & 0 & b_3 & a_4 & b_4 \\ 0 & 0 & 0 & b_4 & a_5 \end{pmatrix}.$$

Lemma Let $T_r = T(1:r, 1:r)$ and $p_r(x) = \det(T_r - xI)$. Then the recursion,

$$p_r(x) = (a_r - x)p_{r-1}(x) - b_{r-1}^2 p_{r-2}(x), \quad p_0(x) = 1,$$

holds.

Remark The polynomial $p_n(x)$ can be evaluated in $\mathcal{O}(n)$ operations.

Suppose $p_n(y)p_n(z) < 0$ and $y < z$ then

```
while |y − z| > ε(|y| + |z|)
    x = (y + z)/z
    if p_n(x)p_n(y) < 0 then
        z=x
    else
        y=x
    end
end
```

Remark This *bisection* procedure is guaranteed to converge. A viable way to compute a couple of eigenvalues.

Diagonal Plus Rank-1

Lemma Suppose $D = \text{diag}(d_1, d_2, \ldots, d_n)$, $d_1 > \cdots > d_n$. Assume $\rho \neq 0$ and that $z \in \mathbb{R}^n$ has no zero components. If

$$(D + \rho z z^T)v = \lambda v, \qquad v \neq 0,$$

then $z^T v \neq 0$ and $D - \lambda I$ is non-singular.

Remark This can be the basis of a recursive algorithm since an update $A^{(k)} := A^{(k-1)} + zz^T$ can split a tridiagonal matrix into two tridiagonal blocks.

Theorem *(Interlacing)* Suppose $B = A + \tau c c^T$, where $A \in \mathbb{R}^{n\times n}$ is symmetric and $\|c\|_2 = 1$. If $\tau > 0$ then

$$\lambda_i(A) \le \lambda_i(B) \le \lambda_{i-1}(A),$$

while if $\tau < 0$ then

$$\lambda_{i+1}(A) \le \lambda_i(B) \le \lambda_i(A).$$

Remark There are many interlacing theorems.

Theorem Suppose $D = \text{diag}(d_1, d_2, \ldots, d_n)$, $d_1 > \cdots > d_n$. Assume $\rho \neq 0$ and that $z \in \mathbb{R}^n$ has no zero components. If V is orthogonal such that

$$V^T(D + \rho zz^T)V = \text{diag}(\lambda_1, \ldots, \lambda_n),$$

with $\lambda_1 \geq \ldots \lambda_n$ and $V = (v_1, \ldots, v_n)$ then

a) The λ_i are the n zeros of $f(\lambda) = 1 + \rho z^T(D - \lambda I)^{-1}z$.

b) The eigenvector v_i is a multiple of $(D - \lambda_i I)^{-1}z$.

Remark To find V we solve $f(z) = 0$, using e.g. Newtons Method, and find v_i by normalizing $(D - \lambda_i)^{-1}z$. Interlacing Theorem places one root in each interval (d_{i-1}, d_i).

How to take advantage of this?

A Divide and Conquer Method

Lemma Let T be symmetric and tridiagonal. There is a c, $\|c\|_2 = 1$, such that

$$T = \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix} + \rho cc^T,$$

where T_1 and T_2 are tridiagonal and ρ is a scalar.

Remark Given two Schur decompositions $Q_1^T T_1 Q_1 = D_1$ and $Q_2^T T_2 Q_2 = D_2$ we can combine

$$U^T T U = D + \rho zz^T, \quad U = \text{diag}(Q_1, Q_2), \quad z = U^T c.$$

Excellent for parallel implementation!

The singular value decomposition

Proposition Every matrix $A \in \mathbb{R}^{m \times n}$ has a decomposition

$$A = U \Sigma V^T,$$

where U and V are orthogonal and $\Sigma \in \mathbb{R}^{m \times n}$ is *diagonal* with diagonal elements $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(n,m)} \geq 0$.

Remark The diagonal elements $\{\sigma_i\}$ are called *singular values* and the columns $\{u_i\}$ of U and the columns $\{v_i\}$ of V are called right and left singular vectors.

Example Compute the SVD in Matlab by

```
>> A=[1 -2 3 ; -2 3 1 ; 2 -4 6 ; -1 2 -3];
>> [U,S,V]=svd(A); U , S
  U =
   -0.4025  0.0684  0.9129 -0.0000
    0.1675  0.9859 -0.0000  0.0000
   -0.8050  0.1368 -0.3651  0.4472
    0.4025 -0.0684  0.1826  0.8944

  S =
    9.2780       0       0
         0  3.4524       0
         0       0  0.0000
         0       0       0
```

Remark The matrix A has rank 2. V is 3×3 orthogonal.

Example Let $A \in \mathbb{R}^{4\times 3}$. Then

$$A = \begin{pmatrix} u_1 & u_2 & u_3 & u_4 \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} v_1 & v_2 & v_3 \end{pmatrix}^T.$$

Remark The vectors $\{u_i\}$ are a basis for $\mathbb{R}^4$ and the vectors $\{v_i\}$ are a basis for $\mathbb{R}^3$

A matrix $A \in \mathbb{R}^{m\times n}$ represents a *linear mapping*

$$A : \mathbb{R}^n \mapsto \mathbb{R}^m.$$

Remark If $U = (u_1, \ldots, u_m) \in \mathbb{R}^{m\times m}$ is *orthogonal* then the set of vectors $\{u_i\}$ form an orthogonal basis for $\mathbb{R}^m$.

Observation In the basis U, V the linear mapping is represented by the diagonal matrix D.

Lemma Let $A \in \mathbb{R}^{m\times n}$ and $A = U\Sigma V^T$. it holds that

$$Av_i = \sigma_i u_i \text{ and} A^T u_i = \sigma_i v_i, i = 1, 2, \ldots, \min(m, n).$$

Lemma Let $A \in \mathbb{R}^{m\times n}$ and $A = U\Sigma V^T$. We can write

$$A = \sum_{i=1}^{\min(m,n)} \sigma_i u_i v_i^T.$$

Linear Systems of Equations

Lemma Let $A \in \mathbb{R}^{n\times n}$ be non-singular and $A = U\Sigma V^T$. Then the solution to the linear system $Ax = b$ is given by

$$x = V\Sigma^{-1}U^T b = \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i.$$

Remarks The solution exists, i.e. A is non-singular, if $\sigma_n > 0$. If σ_n is very small the system is *Ill–conditioned*.

More expensive compared to using the LU factorization. Reveals linear dependencies among the columns of A.

Norms and the Condition Number

Recall If U is orthogonal and x is a vector then $\|Ux\|_2 = \|x\|_2$.

Lemma The norm is $\|A\|_2 = \sigma_1$.

Corollary The condition number is $\kappa_2(A) = \frac{\sigma_1}{\sigma_n}$.

Remark Previously we used $\|A\|_2 = (\lambda_{\max}(A^TA))^{1/2}$. Since $A^TA = U\Sigma^T\Sigma U^T$ we get $\lambda_i(A^TA) = \sigma_i^2$.