HW4 Theory

Monday, March 1, 2021 1:54 PM

1) **1.** The Schur decomposition theorem states that every square matrix $\mathbf{A} \in \mathbb{C}^{m\times m}$ has a Schur decomposition, $\mathbf{A} = \mathbf{QUQ}^*$, where Q is unitary and T is upper-triangular. Use this theorem to prove that, for an arbitrary norm $||\cdot||$,

$$\lim_{n\to\infty} ||\mathbf{A}^n|| = 0 \iff \rho(\mathbf{A}) < 1. \qquad (4)$$

Proof:

($\Rightarrow$) Assume $\lim_{n\to\infty} \|A^n\|_2 = 0$. Then, for Q, U as above,

$$A = QUQ^* \Rightarrow \lim_{n\to\infty} \|A^n\|_2 = \lim_{n\to\infty} \|(QUQ^T)^n\|_2$$

$$= \lim_{n\to\infty} \|(QUQ^T)(QUQ^T)\ldots(QUQ^T)\|_2$$

$$= \lim_{n\to\infty} \|QU^nQ^T\|_2 = \lim_{n\to\infty} \|U^n\|_2 = 0.$$

Hence, as $n\to\infty$, U^n becomes the $0_{m\times m}$ matrix. Since the product of an upper-triangular matrix with itself is indeed upper-triangular with its diagonal entries squared, we know that

$$U_{ii} < 1, \quad i = 1, \ldots, m$$ (otherwise it would not approach the 0 matrix!)

to begin with. Since the eigenvalues of U also lie along its diagonal, we have $\rho(U) < 1$.

Since we decomposed A via the Schur-decomposition, U is similar to A, and hence they share the same spectrum, i.e $\rho(A) < 1$.

($\Leftarrow$) Now assume $\rho(A) < 1$. Then by the

($\Leftarrow$) Now, assume that $\rho(A) < 1$. Then, by the Schur-decomposition, $\exists\, U, Q$, (as above) such that

$$A = QUQ^T, \quad \rho(U) = \rho(A) \Rightarrow \rho(U) < 1.$$

Since U is upper-triangular, its eigenvalues lie along its diagonal. Thus, remembering that

$\rho(U) = \max\{|\lambda_1|, |\lambda_2|, \dots, |\lambda_m|\}$, we know that the diagonal entries of U <u>must</u> be < 1.

Hence, as we repeatedly multiply the upper-triangular U

$$\lim_{n\to\infty} U^n = U \cdot U \cdot U \cdot \dots = 0_{m\times m}.$$

Since $\lim_{n\to\infty} U^n = 0$, we have $\lim_{n\to\infty} \|U^n\|_2 = 0$.

Therefore, for the Q given above, we have

$$\lim_{n\to\infty} \|U^n\|_2 = 0 = \lim_{n\to\infty} \|QU^nQ^T\|_2 = \lim_{n\to\infty} \|\underbrace{(QUQ^T)\cdot \dots \cdot (QUQ^T)}_{n \text{ products}}\|_2$$

$$= \lim_{n\to\infty} \|A^n\|_2$$

Hence, $\lim_{n\to\infty} \|A^n\|_2 = 0$, $\Rightarrow \lim_{n\to\infty} \|A^n\| = 0$ since the norms are equivalent. ∎

2) **2.** **(a)** Let $\mathbf{A} \in \mathbb{C}^{m \times m}$ be tridiagonal and hermitian, with all its sub- and super-diagonal entries nonzero. Prove that the eigenvalues of $\mathbf{A}$ are distinct. (Hint: Show that for any $\lambda \in \mathbb{C}$, $\mathbf{A} - \lambda\mathbf{I}$ has rank at least $m - 1$.)

(b) On the other hand, let $\mathbf{A}$ be upper-Hessenberg, with all its sub-diagonal entries nonzero. Give an example that shows that the eigenvalues of $\mathbf{A}$ are not necessarily distinct.

a) Let A be as above, and λ an eigenvalue of A.

Then, $A - \lambda I =$

$$\begin{bmatrix} a_{11}-\lambda & b_{12} & & & 0 \\ c_{21} & a_{22}-\lambda & b_{23} & & \\ & c_{32} & \ddots & \ddots & \\ & & \ddots & \ddots & b_{m-1,m} \\ 0 & & & c_{m,m-1} & a_{mm}-\lambda \end{bmatrix}.$$

Since $c_{21}, c_{32}, \dots, c_{m,m-1}$ are nonzero, the submatrix

of the matrix $A - \lambda I$

$$\begin{bmatrix} a_{11}-\lambda & b_{12} & & & 0 \\ c_{21} & a_{22}-\lambda & b_{23} & & \\ & c_{32} & \ddots & \ddots & \\ & & \ddots & \ddots & b_{m-1,m} \\ 0 & & & c_{m,m-1} & a_{mm}-\lambda \end{bmatrix}$$

is full rank, since it is upper-triangular. Hence, the matrix $A-\lambda I$ has rank at least $m-1$, since it contains a submatrix of rank $m-1$. To see why that is true, if we consider row reducing the entire matrix $A-\lambda I$, the highlighted submatrix would fully reduce to $I_{(m-1)\times(m-1)}$, and we could simply perform row swaps to align it along the diagonal of $A-\lambda I$.

Since we've shown $\text{rank}(A-\lambda I) \geq m-1$, if λ is an eigenvalue of A, then $\text{rank}(A-\lambda I) = m-1$ (or it would fail to be an eigenvalue). Thus, by the rank-nullity theorem, $\dim(\text{null}(A-\lambda I)) = 1$, i.e. λ has geometric multiplicity 1. Because A is Hermitian, A is diagonalizable. Hence, each eigenvalue of A has

$$\text{geometric mult.} = \text{algebraic mult.}$$

Therefore, the algebraic multiplicity of arbitrary λ is 1. Thus, the eigenvalues of A are distinct.

b) Let A be upper-Hessenberg with $A_{i+1,i} \neq 0$

for $i = 1, \dots, m$.

Consider the 2×2 matrix

$\begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} = A$, which is vacuously upper Hessenberg.

It has eigenvalue $\lambda = 0$ with multiplicity 2,

since $P_A(\lambda) = (-1-\lambda)(1-\lambda) + 1 = \lambda^2$

$\Rightarrow P_A(\lambda) = 0 = \lambda^2 \Rightarrow \lambda = 0.$

3) **3.** Let $\mathbf{A}$ be $m \times n$ and $\mathbf{B}$ be $n \times m$. Show that the matrices $\begin{bmatrix} \mathbf{AB} & 0 \\ B & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & 0 \\ \mathbf{B} & \mathbf{BA} \end{bmatrix}$ have the same eigenvalues.

$$X_1 = \begin{bmatrix} AB & 0 \\ B & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ B & BA \end{bmatrix} = X_2$$

$$\begin{array}{cc} & m+n \\ m+n & \begin{bmatrix} m \times m & m \times n \\ n \times m & m \times m \end{bmatrix} \end{array}, \quad \begin{array}{cc} m+n & \\ \begin{bmatrix} m \times n & m \times m \\ n \times m & n \times n \end{bmatrix} & m+n \end{array}$$

$$\rightarrow \det(X_1 - \lambda I) = \det\left(\begin{bmatrix} AB - \lambda I & 0 \\ B - \lambda I & -\lambda I \end{bmatrix}\right)$$

$$= \det(AB - \lambda I) \cdot \det(-\lambda I) \qquad (*)$$

$$\rightarrow \det(X_2 - \lambda I) = \det\left(\begin{bmatrix} -\lambda I & 0 \\ B & BA-\lambda I \end{bmatrix}\right)$$

$$= \det(BA - \lambda I) \cdot \det(-\lambda I). \quad (**)$$

We have proved previously that AB and BA have the same spectrum.

The roots of the expression

$\det(AB - \lambda I)$ are the nonzero eigenvalues of AB, and the roots of the expression $\det(BA - \lambda I)$ are the nonzero eigenvalues of BA.

The extra factor of $\det(-\lambda I)$ in $(*)$, $(**)$ add nothing but zero eigenvalues.

Hence X_1 and X_2 must have the same set of eigenvalues (although multiplicities may vary).

4)

4. Show that the Gerschgorin theorem also holds with the bounds r_i which are given by the partial column sums (instead of the partial row sums):

$$r_i = \sum_{i=1, i\neq j}^{m} |a_{i,j}|, \quad i = 1, \ldots, m. \tag{5}$$

Proof: We simply apply the theorem to A^T.

Let $A = \{a_{ij}\}$ be an $n \times n$ matrix, so

$A^T = \{a_{ji}\}$

The Gershgorin theorem states that every eigenvalue of matrix A^T lies within at least one of the Gershgorin discs $D(a_{ii}, R_i)$,

where $R_i = \sum_{j \neq i}^{n} |a_{ji}|$ ← row sum of A^T which is the column sum of A.

Recall A and A^T have the same spectrum (HW 1, 2c).
Thus, we can immediately conclude the theorem holds for partial column sums as well, since an eigenvalue λ of A must lie in the disc with radius partial row sum of A^T = partial column sum of A.

$\blacksquare$

5)

5. Use the Gerschgorin theorem to show that the matrix

$$\mathbf{A} = \begin{bmatrix} 1.0 & 0.3 & 0.1 & 0.4 \\ 0.0 & 2.0 & 0.0 & 0.1 \\ 0.0 & 0.4 & 3.0 & 0.0 \\ 0.1 & 0.0 & 0.0 & 4.0 \end{bmatrix} \tag{6}$$

has exactly one eigenvalue in each of the four circles

$$|z - k| \leq 0.1, \quad k = 1, 2, 3, 4. \tag{7}$$

Every eigenvalue of A will lie in at least one disc $D(a_{ii}, R_i)$, where $R_i = \sum_{j \neq i} |a_{ij}|$ is the radius. Note the Gershgorin circle theorem also holds for the column sums, allowing us to pick a potentially smaller radius.

For a_{11}: $a_{11} = 1.0$, $R_1 = 0.8$ or $R_1 = 0.1$ ← pick this radius because it is smaller.

For a_{22}: $a_{22} = 2.0$, $R_2 = 0.1$ or $R_2 = 0.7$

For a_{33}: $a_{33} = 3.0$, $R_3 = 0.4$ or $R_3 = 0.1$

For a_{44}: $a_{44} = 4.0$, $R_4 = 0.1$ or $R_4 = 0.5$.

Hence, we've shown that centered at each diagonal entry, there is a Gershgorin disc of radius 0.1. Since all disc centers are 1.0 apart, they never intersect. Thus, we know that A has exactly one eigenvalue in each disc.

6)

6. Let $\mathbf{A} \in \mathbb{R}^{m \times m}$ be strictly diagonally dominant matrix, namely,

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}|, \quad 1 \leq i \leq m. \tag{8}$$

Show that $\mathbf{A}$ is nonsingular. Also, find a lower bound for $||\mathbf{A}^{-1}||_2$ using the Gerschgorin theorem.

Proof: Let A_m be as above.

Since $|a_{ii}| > \sum_{j\neq i} |a_{ij}|$, $1 \leq i \leq m$, we know that $|a_{ii}| \neq 0$ from the strict inequality.

Hence, by the Gershgorin circle theorem, we know that every eigenvalue λ lies within at least one Gershgorin disc of radius $\sum_{j\neq i}^{m} |a_{ij}| = R_i$.

Therefore, since $R_i < |a_{ii}| \neq 0$, every disc centered at a_{ii}, $1 \leq i \leq m$ does not include 0.

Since all eigenvalues of A lie within at least one disc, and not one disc contains 0, A cannot have $\lambda = 0$ as an eigenvalue, i.e. A is full-rank.

Thus A is nonsingular.

From HW1 Problem 6, we know that

$\|A^{-1}\| \geq \rho(A^{-1})$, where $\rho(A^{-1}) = \lambda_{max}(A^{-1}) = \dfrac{1}{\lambda_{min}(A)}$.

7. Let $\mathbf{A} \in \mathbb{R}^{m\times m}$ be non-defective (so that there exist eigenvectors $\mathbf{v}_i$ corresponding to λ_i that form a basis) and let $\{\lambda_i\}_{i=1}^m$ be its eigenvalues. Assume $\lambda_1 = \cdots = \lambda_r$ for some $1 < r < m$ and

7)

$$|\lambda_r| > |\lambda_{r+1}| \geq \cdots \geq |\lambda_m|. \tag{9}$$

Consider the Power Iteration method given by

$$\mathbf{x}^{(k+1)} = \mathbf{A}\mathbf{x}^{(k)}, \quad \mathbf{y}^{(k+1)} = \frac{\mathbf{x}^{(k+1)}}{||\mathbf{x}^{(k+1)}||}, \tag{10}$$

with any arbitrary norm $||\cdot||$. Show that, for almost all possible choices of an initial vector $\mathbf{x}^{(0)}$, $\mathbf{y}^{(k+1)}$ will converge to an eigenvector associated to λ_1. For which values of $\mathbf{x}_{(0)}$ does the method *not* converge to such an eigenvector?

<u>Proof</u>: Assume A is as above. Then, since $\{V_i\}_{i=1,\dots,m}$ form a basis, we can write

$$x = \sum_{i=1}^{m} \alpha_i V_i, \quad \alpha_i \in \mathbb{R}.$$ Then, we have

$$Ax = \sum_{i=1}^{m} \lambda_i \alpha_i V_i \Rightarrow A^n x = \sum_{i=1}^{m} \lambda_i^n \alpha_i V_i.$$

Then, factoring out λ_r^n,

$$A^n x = \lambda_r^n \sum_{i=1}^{m} \left(\frac{\lambda_i}{\lambda_r}\right)^n \alpha_i v_i.$$ Since λ_r is the largest eigenvalue of A, we know $\left|\frac{\lambda_i}{\lambda_r}\right| < 1$ for $i > r$. NOT DONE, unsure how to proceed.

8) **8.** Let $\mathbf{A} \in \mathbb{R}^{m\times m}$ be a non-defective matrix with its eigenvalues $\{\lambda_i\}_{i=1}^m$ and its singular values $\{\sigma_i\}_{i=1}^m$, satisfying

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_m|, \tag{11}$$
$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m. \tag{12}$$

Let $\rho(\mathbf{A})$ be the spectral radius of $\mathbf{A}$ and $\mathrm{cond}(\mathbf{A}) = ||\mathbf{A}||_2||\mathbf{A}^{-1}||_2$ be the condition number of $\mathbf{A}$. Show that:

(a) $\sigma_i = |\lambda_i|$, $1 \leq i \leq m$, if $\mathbf{A}$ is normal, i.e., $\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T$.

(b) $||\mathbf{A}||_2 = |\lambda_1| = \rho(\mathbf{A})$.

a) Assume A is normal. Then, the spectral theorem for normal matrices gives us that

$A = PDP^*$, where D is diagonal and $D_{ii} = \lambda_i(A)$ for $i = 1, \ldots, m$ and P is orthogonal.

Since the eigenvalues of A aren't necessarily real, we have that

$$A^*A = (PD^*P^*)(PDP^*) = PD^*DP^*.$$

Note then that $(D^*D)_{ii} = |\lambda_i|^2$. Thus,

$$\sigma_i = \sqrt{\lambda_i(PD^*DP^*)}$$

$= \sqrt{\lambda_i(D^*D)}$ since P is orthogonal (or unitary here) and hence invertible.

$= \sqrt{|\lambda_i|^2} = |\lambda_i|$ $\blacksquare$

b) Proof:

We want to show $\|A\|_2 = |\lambda_1| = \rho(A)$.

Assuming A is normal, then from (a) we know $\sigma_i = |\lambda_i|$, $1 \le i \le m$.

By definition,

$\|A\|_2 = \sigma_1$, the largest singular value of A.

Thus by (a),

$\|A\|_2 = \sigma_1 = |\lambda_1|$. $\blacksquare$

9)

9. Let $\mathbf{A} \in \mathbb{R}^{m\times m}$ be symmetric and positive definite. Suppose that you have found the following algorithm that is known to produce a sequence of $\mathbf{A}_i$. The algorithm claims that $\mathbf{A}_i$ converges to a diagonal matrix $\Lambda = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_m)$, where $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_m > 0$ as $i \to \infty$. In each iteration, the algorithm uses the Cholesky decomposition to factorize $\mathbf{A}_i$ to an upper triangular matrix $\mathbf{U}_i$ whose diagonal elements are nonzero. The algorithm proceeds as follows:

Algorithm:

$\mathbf{A}_0 = \mathbf{A}$
for $i = 1, \cdots$
$\mathbf{U}_i^T\mathbf{U}_i = \mathbf{A}_{i-1}$ `[Cholesky Decomposition]`
$\mathbf{A}_i = \mathbf{U}_i\mathbf{U}_i^T$
endfor

Prove that the iteration is well-defined by showing the following steps:

(a) Show that $\mathbf{A}_i$ is also symmetric and positive definite to justify the use of Cholesky.

$\mathbf{A}_0 = \mathbf{A}$
for $i = 1, \cdots$
$\mathbf{U}_i^T \mathbf{U}_i = \mathbf{A}_{i-1}$ **[Cholesky Decomposition]**
$\mathbf{A}_i = \mathbf{U}_i \mathbf{U}_i^T$
endfor

Prove that the iteration is well-defined by showing the following steps:

(a) Show that $\mathbf{A}_i$ is also symmetric and positive definite to justify the use of Cholesky.

(b) Show that $\mathbf{A}_i$ is similar to $\mathbf{A}$ (i.e., $\mathbf{A}_i = \mathbf{B}^{-1}\mathbf{A}\mathbf{B}$ for some non-singular matrix $\mathbf{B}$).

(c) Explain why you can use this iteration method as a valid eigenvalue revealing algorithm for $\mathbf{A}$.

a) It suffices to show that (a) holds for $i=1$.

Proof: We have $A_0 = A$, hence A_0 is positive definite. Then,

$$U_1^T U_1 = A_0 \rightsquigarrow \text{Cholesky decompose } A_0.$$

and

$$A_1 = U_1 U_1^T.$$

It remains to show that A_1 is symmetric positive definite.

Clearly, since $A_1^T = (U_1^T)^T (U_1) = U_1 U_1^T = A_1$, we know A_1 is symmetric. Now, we want to show

$$x^T A_1 x > 0 \text{ for all } x \neq 0.$$

$$\rightarrow \quad x^T A_1 x = x^T U_1 U_1^T x = (U_1^T x)^T (U_1^T x) = \|U_1^T x\|_2^2.$$

We know $\|U_1^T x\| > 0$, since U_1 is full-rank

We know $\|U_1^T x\| > 0$, since U_1 is full-rank $\searrow U_{ii} > 0$
(and thus U_1^T is as well), implying $U_1^T x = 0$ iff $x = 0$.
Thus A_1 is symmetric positive definite.
This proves (a), since for $i=2$, we know A_1 is sym. pos. def., and hence we can perform the Cholesky decomposition on it and follow the same steps as above, producing sym. pos. def. A_2, and so on. $\blacksquare$

b) Proof:

We have that

$$A_{i-1} = U_i^T U_i$$

$$A_i = U_i U_i^T.$$

From (a), A_k is symmetric positive definite and hence U_k is full-rank (thus invertible) for all $k \in \mathbb{N}$.

Hence, A_{i-1} is similar to A_i since we can write

$$A_{i-1} = U_i^T U_i = U_i^{-1}(U_i U_i^T)U_i$$
$$= U_i^{-1} A_i U_i \text{ for any } i.$$

$\Rightarrow$ A_i is similar to A_{i+1}, hence we can conclude $A = A_0$ is similar to any A_i.

$\blacksquare$

c) Assume that A_i converges to a diagonal matrix. From part (b) we know A and A_i are similar, hence have the same spectrum. Thus, the diagonal entries of A_i must be the eigenvalues of A.