

Notes on Chapter 5

1. If H is Hermitian ($H = H^*$), all eigenvalues of H are real.

Proof: if $H z = \lambda z$, then

$$\lambda z^* z = z^*(\lambda z) = z^*(H z) = (H^* z)^* z = (H z)^* z = (\lambda z)^* z = \bar{\lambda} z^* z$$

Since $z^* z = \|z\|^2 \neq 0$, then $\lambda = \bar{\lambda}$ so λ is real.

2. If H is Hermitian, we prove that H is "unitarily similar" (unitarily equivalent) to a diagonal matrix, that is, that $H = SDS^{-1}$, where D is diagonal and S is unitary. Note that since $HS = SD$, this means that H has an orthonormal basis of eigenvectors. The proof is by induction. Obviously it is true for any 1 by 1 matrix, just take $S=I$. Assume the claim is true for every $n-1$ by $n-1$ matrix. Then let H be an n by n Hermitian matrix; and let λ_1 be an eigenvalue, with eigenvector u_1 , of norm (2-norm) one. Then put u_1 in the first column of S , and pick the other columns of S to complete an "orthonormal" basis for C^n , so that S is unitary, and then $HS = SB$, or $H = SBS^{-1}$ where B has the form:

$$\begin{bmatrix} \lambda_1 & w^T \\ 0 & B_1 \end{bmatrix}$$

Now since H is Hermitian, $B = S^{-1}HS = S^*HS$ is also Hermitian, because $B^* = (S^*HS)^* = S^*H^*S = S^*HS = B$. Thus the first row of B is just the conjugate of the first column, which means $w = 0$, and B_1 must be a Hermitian $n-1$ by $n-1$ matrix. Thus by assumption, $B_1 = P_1 D_1 P_1^{-1} = P_1 D_1 P_1^*$ where D_1 is diagonal and P_1 is unitary. Then:

$$B = \begin{bmatrix} \lambda_1 & 0 \\ 0 & B_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & P_1 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & D_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & P_1^* \end{bmatrix}$$

or $B = PDP^*$, where P is unitary.

Now since $H = SBS^* = S(PDP^*)S^* = (SP)D(SP)^*$, H is unitarily similar to a diagonal matrix (note that SP is unitary).

3. A "positive definite" matrix is a Hermitian matrix whose eigenvalues are all positive. An equivalent definition is, it is a Hermitian matrix such that $z^* H z > 0$ for all $z \neq 0$. Proof that the two definitions are equivalent: If

all eigenvalues are positive, $H = S^*DS$ for a unitary matrix S and diagonal matrix D with all diagonal entries positive. Then

$$z^*Hz = z^*(S^*DS)z = (Sz)^*D(Sz) = \sum_{i=1}^N \lambda_i |s_i|^2 > 0$$

where the s_i are the components of the (non-zero) vector Sz . On the other hand, if $z^*Hz > 0$ for all nonzero vectors z , all eigenvalues of H are positive, because if $H z = \lambda z$, $0 < z^*Hz = \lambda z^*z = \lambda \|z\|_2^2$ and so $\lambda > 0$.

Note that A^*A for any A (even non-square matrices) is Hermitian, and at least "positive semi-definite", which means all eigenvalues are greater than or equal to 0. Proof: if $A^*Az = \lambda z$, then

$$\lambda \|z\|^2 = \lambda z^*z = z^*(\lambda z) = z^*(A^*Az) = (Az)^*Az = \|Az\|^2 \geq 0, \text{ so } \lambda \geq 0.$$

4. The Power Method

Suppose A is diagonalizable (eg, unitary, Hermitian, all distinct eigenvalues,...) and it has one eigenvalue λ_1 which is greater in absolute value than all others. Then if x_0 is chosen "at random" and the power method is defined by $x_{k+1} = Ax_k$, then $\lambda_1^{-k}x_k$ converges to an eigenvector corresponding to λ_1 , and the quotient $\frac{\langle x_{k+1}, x_k \rangle}{\langle x_k, x_k \rangle}$ converges to λ_1 , with probability one. Proof:

$$\lambda_1^{-k}x_k = \lambda_1^{-k}A^kx_0 = S(\lambda_1^{-k}D^k)S^{-1}x_0$$

where D is a diagonal matrix with the eigenvalues of A on the diagonal, and $D_{ii} = \lambda_i$ for $i = 1, \dots, m_1$, where m_1 is the multiplicity of λ_1 . Since λ_1 is larger in absolute value than the other eigenvalues, as k goes to infinity, $\lambda_1^{-k}D^k$ converges to a diagonal matrix E , with $E_{ii} = 1$ for $i = 1, \dots, m_1$ and $E_{ii} = 0$ otherwise. So $\lambda_1^{-k}x_k$ converges to $f = SES^{-1}x_0$. To verify that f is an eigenvector, we compute $Af = SDS^{-1}SES^{-1}x_0 = SDES^{-1}x_0 = \lambda_1SES^{-1}x_0 = \lambda_1f$. Note that we have used $DE = \lambda_1E$, which can be verified by writing out the diagonal matrices D and E . Now SES^{-1} is not the zero matrix, because if it were, E would be the zero matrix and it isn't (quite). Therefore the null space of SES^{-1} is of dimension $N-1$ or less, and so the probability that f is not the zero vector is one, if x_0 is chosen randomly. Finally, $\frac{\langle x_{k+1}, x_k \rangle}{\langle x_k, x_k \rangle} = \lambda_1 \frac{\langle \lambda_1^{-(k+1)}x_{k+1}, \lambda_1^{-k}x_k \rangle}{\langle \lambda_1^{-k}x_k, \lambda_1^{-k}x_k \rangle}$ which converges to

$$\lambda_1 \frac{\langle f, f \rangle}{\langle f, f \rangle} = \lambda_1.$$

- a. The power method still works even if A is not diagonalizable, but the proof is more difficult. It will not work if there is more than one eigenvalue tied for largest in absolute value; to find the largest eigenvalue,

there has to be a largest eigenvalue. Note that a largest eigenvalue of multiplicity greater than 1 is ok, but two different eigenvalues tied for largest is not ok.

- b. Note that the speed of convergence is faster if the second largest eigenvalue is much smaller than the largest, because then $\lambda_1^{-k} D^k$ converges to E faster.
- c. The power method only finds the largest eigenvalue. However, since the eigenvalues of $B = (A - pI)^{-1}$ are $(\lambda_i - p)^{-1}$, we can apply the power method to find the largest eigenvalue of B, which will be $(\lambda_p - p)^{-1}$, where λ_p is the eigenvalue of A closest to p. So we can choose a value for p and "go fishing" for the eigenvalue of A closest to p. (Choose p=0 if you want the smallest eigenvalue of A.)

5. The Singular Value Decomposition

The "singular value decomposition" of an arbitrary M by N matrix is $A = UDV^*$, where U is an M by M unitary matrix, V is an N by N unitary matrix, and D is an M by N diagonal matrix. The diagonal elements of D are called the singular values of A. First, let's suppose A is a nonsingular square (N by N) matrix. Even though A may be nonsymmetric, A^*A is Hermitian and positive definite, so there exists a unitary matrix V such that $V^*A^*AV = D^2$, where D^2 is diagonal with the (real and positive) eigenvalues of A^*A . Then we set $U = AVD^{-1}$ and $A = UDV^*$ is a singular value decomposition, because $UDV^* = AVD^{-1}DV^* = AVV^* = A$, and $U^*U = (AVD^{-1})^*AVD^{-1} = D^{-1}(V^*A^*AV)D^{-1} = D^{-1}D^2D^{-1} = I$, so U is unitary. Notice that the singular values are just the square roots of the eigenvalues of A^*A , which are positive if A is nonsingular, and nonnegative for any A. The largest singular value of A is equal to $\|A\|_2$, where this is the matrix norm subordinate to the L_2 vector norm. Details on how to find the singular value decomposition for general (rectangular) A can be found in my Computational Linear Algebra text, in problem 3.17.

The singular value decomposition has a number of uses, here is one. If $A = UDV^*$, and u_i is the i^{th} column of U, v_i^* is the i^{th} row of V^* , and D_{ii} is the i^{th} element of the diagonal matrix D, we can show that $A = \sum_i D_{ii}u_iv_i^*$. If A is a large matrix with many zero or nearly zero singular values D_{ii} , we can throw away the corresponding small terms in this series, leaving us with a more compact approximate representation of A. If there are only L terms

left after we discard the smaller terms, then we have an approximation of the M by N matrix A which requires only $L(M+N)$ words of memory, which may be small compared to MN . See p195 of the Allen text for an example of how the singular value decomposition can be used to compress a bit matrix picture, without losing too much clarity.