Chapter 5

Hermitian Theory

Hermitian matrices form one of the most useful classes of square matrices. They occur naturally in a variety of applications from the solution of partial differential equations to signal and image processing. Fortunately, they possess the most desirable of matrix properties and present the user with a relative ease of computation. There are several very powerful facts about Hermitian matrices that have found universal application. First the spectrum of Hermitian matrices is real. Second, Hermitian matrices have a complete set of orthogonal eigenvectors, which makes them diagonalizable. Third, these facts give a spectral representation for Hermitian matrices and a corresponding method to approximate them by matrices of less rank.

5.1 Diagonalizability of Hermitian Matrices

Let's begin by recalling the basic definition.

Definition 5.1.1. Let $A \in M_n(\mathbb{C})$. We say that A is *Hermitian* if $A = A^*$, where $A^* = \bar{A}^T$. A^* is called the adjoint of A. This, of course, is in conflict with the other definition of adjoint, which is given in terms of minors.

Recall the following facts and definitions about subspaces of $\mathbb{C}_n$:

- If U, V are subspaces of $\mathbb{C}_n$, we define the *direct sum* of U and V by $U \oplus V = \{u + v \mid u \in U, \ v \in V\}$.

- If U, V are subspaces of $\mathbb{C}_n$, we say U and V are orthogonal if $\langle u, v \rangle = 0$ for every $u \in U$ and $v \in V$. In this case we write $U \perp V$.

For example, a natural way to obtain orthogonal subspaces is from orthonormal bases. Suppose that $\{u_1, \ldots, u_n\}$ is an orthonormal basis of $\mathbb{C}_n$. Let

the integers $\{1, \ldots, n\}$ be divided into two (disjoint) subsets J_1 and J_2. Now define

$$\begin{aligned} U_1 &= \mathfrak{S}\{u_i \mid i \in J_1\} \\ U_2 &= \mathfrak{S}\{u_i \mid i \in J_2\} \end{aligned}$$

Then U_1 and U_2 are orthogonal, i.e. $U_1 \perp U_2$, and

$$U_1 \oplus U_2 = \mathbb{C}_n$$

$$\langle Au, x \rangle = \langle u, Ax \rangle = \lambda \langle u, x \rangle = 0$$

Our main result is that Hermitian matrices are diagonalizable. To prove it, we reveal other interesting and important properties of Hermitian matrices. For example, consider the following.

Theorem 5.1.1. *Let $A \in M_n(\mathbb{C})$ be Hermitian. Then the spectrum of $A, \sigma(A)$, is real.*

Proof. Let $\lambda \in \sigma(A)$ with corresponding eigenvector $x \in \mathbb{C}_n$. Then

$$\langle Ax, x \rangle = \langle x, Ax \rangle = \langle x, \lambda x \rangle = \bar{\lambda}\langle x, x \rangle$$
$$\|$$
$$\langle \lambda x, x \rangle$$
$$\|$$
$$\lambda \langle x, x \rangle.$$

Since we know $\|x\|^2 = \langle x, x \rangle \neq 0$, it follows that $\lambda = \bar{\lambda}$, which is to say that λ is real. $\qquad\square$

Theorem 5.1.2. *Let $A \in M_n(\mathbb{C})$ be Hermitian and suppose that λ and μ are different eigenvalues with corresponding eigenvectors x and y. Then $x \perp y$ (i.e. $\langle x, y \rangle = 0$).*

Proof. We know $Ax = \lambda x$ and $Ay = \mu y$. Now compute

$$\langle Ax, y \rangle = \langle x, A^*y \rangle = \langle x, Ay \rangle = \mu\langle x, y \rangle$$
$$\|$$
$$\lambda\langle x, y \rangle.$$

If $\langle x, y \rangle \neq 0$, the equality above yields a contradiction and the result is proved. $\qquad\square$

Remark 5.1.1. This result also follows from the previously proved result about the orthogonality of left and right eigenvectors pertaining to different eigenvalues.

Theorem 5.1.3. *Let $A \in M_n(\mathbb{C})$ be Hermitian, and let λ be an eigenvalue of A. Then the algebraic and geometric multiplicities of λ are equal. In symbols,*

$$m_a(\lambda) = m_g(\lambda).$$

Proof. We prove this result by reconsideration of our main result on triangularization of A by a similarity transformation. Let $x \in \mathbb{C}_n$ be an eigenvector of A pertaining to λ. So, $Ax = \lambda x$. Let $u_2, \ldots, u_n \subset \mathbb{C}_n$ be a set of vectors orthogonal to x, so that $\{x, u_2, \ldots, u_n\}$ is a basis of $\mathbb{C}_n$. Indeed, it is an orthogonal basis, and by normalizing the vectors it becomes an orthonormal basis. We claim that $U_2 = \mathfrak{S}(u_2, \ldots, u_n)$, the span of $\{u_2, \ldots, u_n\}$ is invariant under A. To see this, suppose

$$u = \sum_{j=2}^{n} c_j u_j \in U_2$$

and

$$Au = v + ax$$

where $v \in U_2$ and $a \neq 0$. Then, on the one hand

$$\langle Au, x \rangle = \langle u, Ax \rangle = \lambda \langle u, x \rangle = 0$$

On the other hand

$$\langle Au, x \rangle = \langle v + ax, x \rangle = a \langle x, x \rangle = a \, \|x\|^2 \neq 0$$

This contraction establishes that the span of $\{u_2, \ldots, u_n\}$ is invariant under A.

Let $U \oplus V = \mathbb{C}^n$ be invariant subspaces with $U \perp V$. Suppose that U_B and V_B are orthonormal bases of the orthogonal subspaces U and V. Define the matrix $P \in M_n(C)$ by taking for its columns first the basis vectors U_B and and then the basis vectors V_B. Let us write $P = [U_B, V_B]$ (with only a small abuse of notation). Then, since A is Hermitian,

$$B = P^{-1}AP$$

$$= \left[\begin{array}{c|c} A_u & 0 \\ \hline 0 & A_v \end{array} \right]$$

The whole process can be carried out exactly $m_a(\lambda)$ times, each time generating a new orthogonal eigenvector pertaining to λ. This establishes that $m_g(\lambda) = m_a(\lambda)$. A formal induction could have been given. □

Remark 5.1.2. Note how we applied orthogonality and invariance to force the triangular matrix of the previous result to become diagonal. This is what permitted the successive extraction of eigenvectors. Indeed, if for any eigenvector x the subspace of $\mathbb{C}_n$ orthogonal to x is invariant, we could have carried out the same steps as above.

We are now in a position to state our main result, whose proof is implicit in the three lemmas above.

Theorem 5.1.4. *Let $A \in M_n(\mathbb{C})$ be Hermitian. Then A is diagonalizable. The matrix P for which $P^{-1}AP$ is diagonal can be taken to be orthogonal. Finally, if $\{\lambda_1, \ldots, \lambda_n\}$ and $\{u1, \ldots, u_n\}$ denote eigenvalues and pertaining orthonormal eigenvectors for A, then A admits the spectral representation $A = \sum_{j=1}^{n} \lambda_j u_j u_j^T$.*

Corollary 5.1.1. *Let $A \in M_n(\mathbb{C})$ be Hermitian.*

   (i) A has n linearly independent and orthogonal eigenvectors.

   (ii) A is unitarily equivalent to a diagonal matrix.

   (iii) If $A, B \in M_n$ are unitarily equivalent, then A is Hermitian if and only if B is Hermitian.

Note that in part (iii) above, the condition of unitary equivalence cannot be replaced by just similarity. (Why?)

Theorem 5.1.5. *If $A, B \in M_n(\mathbb{C})$ and $A \sim B$ with S as the similarity transformation matrix, $B = S^{-1}AS$. If $Ax = \lambda x$ and $y = S^{-1}x$, then $By = \lambda y$.*

If matrices are similar so also are their eigenstructures. It should establish the very closeness that similarity implies. Later as we consider decomposition theorems, we will see even more remarkable consequences.

Though we have as yet no method of determining the eigenvalues of a matrix beyond factoring the characteristic polynomial, it is instructive to see how their existence impacts the fundamental problem of solving $Ax = b$. Suppose that A is Hermitian with eigenvalues $\lambda_1, \ldots, \lambda_n$, counted according to multiplicity and with orthonormal eigenvectors $\{u_1, \ldots, u_n\}$. Consider

the following solution method for the system $Ax = b$. Since the span of the eigenvectors is C_n then

$$b = \sum_{i=1}^{n} b_i u_i$$

where as we know by the orthonormality of the vectors $\{u_1, \ldots, u_n\}$ that $b_i = \langle b, u_i \rangle$. We can also write $x = \sum_{i=1}^{n} x_i u_i$. Then, the system becomes

$$\begin{aligned} Ax &= A \left(\sum_{i=1}^{n} x_i u_i \right) \\ &= \sum_{i=1}^{n} x_i \lambda_i u_i = \sum_{i=1}^{n} b_i u_i \end{aligned}$$

Therefore, the solution is

$$x_i = \frac{b_i}{\lambda_i}, \quad i = 1, \ldots, n$$

Expanding the data vector b in the basis of eigenvectors yields a rapid method to find the solution to the system. Nonetheless, this is not the preferred method for solving linear systems when the coefficient matrix is Hermitian. Finding all the eigenvectors is usually costly, and other ways are available that are more efficient. We will discuss a few of them in in the section and in later chapters.

Approximating Hermitian matrices

With the spectral representation available, we have a tool to approximate the matrix, keeping the "important" part and discarding the less important part. Suppose the eigenvalues are arranged in decending order $|\lambda_1| \geq \cdots \geq |\lambda_n|$. Now approximate A by

$$A_k = \sum_{j=1}^{k} \lambda_j u_j u_j^T \qquad (1)$$

This is an $n \times n$ matrix. The difference $A - A_k = \sum_{j=k+1}^{n} \lambda_j u_j u_j^T$. We can approximate the norm of the difference by

$$(A - A_k) x = \left(\sum_{j=k+1}^{n} \lambda_j u_j u_j^T \right) x = \sum_{j=k+1}^{n} \lambda_j x_j u_j$$

where $x = \sum_{j=1}^{n} x_j u_j$. Assume $\|x\| = 1$. By the Cauchy-Schwartz inequality

$$\|(A - A_k) x\|^2 = \left\| \sum_{j=k+1}^{n} \lambda_j x_j u_j \right\|^2$$

$$\leq \sum_{j=k+1}^{n} |\lambda_j|^2$$

Therefore, $\|(A - A_k)\| \leq \left(\sum_{j=k+1}^{n} |\lambda_j|^2 \right)^{1/2}$. From this we can conclude that if the smaller eigenvalues are sufficiently small, the matrix can be accurately approximated by a matrix of lesser rank.

Example 5.1.1. The matrix

$$A = \begin{bmatrix} 0.5745 & -0.5005 & 0.1005 & 0.0000 \\ -0.5005 & 1.176 & -0.5756 & 0.1005 \\ 0.1005 & -0.5756 & 1.176 & -0.5005 \\ 0.0000 & 0.1005 & -0.5005 & 0.5745 \end{bmatrix}$$

has eigenvalues eigenvectors $\{2.004, 0.9877, \ 0.3219, \ 0.1872\}$ with pertaining eigenvectors

$$u_1 = \begin{bmatrix} 0.2740 \\ -0.6519 \\ 0.6519 \\ -0.2740 \end{bmatrix}, \ u_2 = \begin{bmatrix} 0.4918 \\ -0.5080 \\ -0.5080 \\ 0.4918 \end{bmatrix}, \ u_3 = \begin{bmatrix} 0.6519 \\ 0.2740 \\ -0.2740 \\ -0.6519 \end{bmatrix}, \ u_4 = \begin{bmatrix} 0.5080 \\ 0.4918 \\ 0.4918 \\ 0.5080 \end{bmatrix}$$

respectively.   Neglecting the eigenvectors pertaining to the two smaller eigenvalues A is approximated according as 1 the formula above by

$$A_2 = \sum_{j=1}^{2} \lambda_j u_j u_j^T = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T$$

$$= 2.004 \begin{bmatrix} 0.274 \\ -0.6519 \\ 0.6519 \\ -0.274 \end{bmatrix} \begin{bmatrix} 0.274 \\ -0.6519 \\ 0.6519 \\ -0.274 \end{bmatrix}^T + 0.9877 \begin{bmatrix} 0.4918 \\ -0.508 \\ -0.508 \\ 0.4918 \end{bmatrix} \begin{bmatrix} 0.4918 \\ -0.508 \\ -0.508 \\ 0.4918 \end{bmatrix}^T$$

$$A_2 = \begin{bmatrix} 0.3893 & -0.6047 & 0.1112 & 0.0884 \\ -0.6047 & 1.107 & -0.5968 & 0.1112 \\ 0.1112 & -0.5968 & 1.107 & -0.6047 \\ 0.0884 & 0.1112 & -0.6047 & 0.3893 \end{bmatrix}$$

The difference

$$A - A_2 = \begin{bmatrix} 0.1852 & 0.1042 & -0.0107 & -0.0884 \\ 0.1042 & 0.069 & 0.0212 & -0.0107 \\ -0.0107 & 0.0212 & 0.069 & 0.1042 \\ -0.0884 & -0.0107 & 0.1042 & 0.1852 \end{bmatrix}$$

has 2-norm $\|A - A_2\|_2 = 0.3218$, while the 2-norm $\|A\|_2 = 2.004$. The relative error of approximation is $\frac{\|A - A_2\|_2}{\|A\|_2} = \frac{0.3218}{2.004} = 0.1606$.

To illustrate how this may be used, let us attempt to use A_2 to approximate the solution of $Ax = b$, where $b = [2.606, -4.087, 1.113, 0.346\,4]^T$. First of all the exact solution is $x = [2.223, -2.688, -0.162\,9, 0.931\,2]^T$. Since the matrix A_2 has rank two, it is not solvable for every vector b. We therefore project the vector b into the span of the range of A_2, namely u_1 and u_2. Thus

$$b_2 = \langle b, u_1 \rangle u_1 + \langle b, u_2 \rangle u_2 = [2.555, -4.118, 1.109, 0.358\,4]^T$$

Now solve $A_2 x_2 = b_2$, to obtain $x_2 = [2.023, -2.828, -0.220\,2, 0.927\,4]^T$. The 2-norm of the difference is $\|x - x_2\|_2 = 0.250\,8$. This error, though not extremely small, can be accounted for by the fact that the data vector b has sizable u_3 and u_4 components. That is $\left\|proj_{\mathfrak{S}(u_3,u_4)} b\right\| = 0.250\,8$.

5.2 Finding eigenvectors

Recall that a zero of a polynomial is called **simple** if its multiplicity is one. If the eigenvalues of $A \in M_n(\mathbb{C})$ are distinct and the largest, λ_n, in modulus is simple, then there is an iterative method to find it. Assume

$$\begin{aligned} (1) &\quad |\lambda_n| = \rho(A) \\ (2) &\quad m_a(\lambda_n) = 1. \end{aligned}$$

Moreover, without loss of generality we assume eigenvalues to be ordered $|\lambda_1| \leq \cdots \leq |\lambda_{n-1}| < |\lambda_n| = \rho_n$. Select $x^{(0)} \in \mathbb{C}^{(n)}$. Define

$$x^{(k+1)} = \frac{1}{\|x^{(k)}\|} A x^{(k)}.$$

By scaling we can assume that $\lambda_n = 1$, and that $y^{(1)}, \ldots, y^{(n)}$ are linearly independent eigenvectors of A. So $Ay^{(n)} = y^{(n)}$. We can write

$$x^{(0)} = c_1 y^{(1)} + \cdots + c_n y^{(n)}.$$

Then, except for a scale factor (i.e. the factor $\|x^{(k)}\|^{-1}$)

$$x^{(k)} = c_1 \lambda_1^k y^{(1)} + \cdots + c_n \lambda_n^k y^{(n)}.$$

Since $|\lambda_j| < 1$, $j = 1, 2, \ldots, n-1$, we have that $|\lambda_j^k| \to 0$ if $j = 1, 2, \ldots, n-1$. Therefore, the limit of $x^{(k)}$ approaches a multiple of $y^{(n)}$.

This gives the following result.

Theorem 5.2.1. *Let $A \in M_n(\mathbb{C})$ have n distinct eigenvalues and assume the eigenvalue λ_n with modulus $\rho(A)$ is simple. If $x^{(0)}$ is not orthogonal to the eigenvector $y^{(n)}$ pertaining to λ_n, then the sequence of vectors defined by*

$$x^{(k+1)} = \frac{1}{\|x^{(k)}\|} A x^{(k)}$$

converges to a multiple of $y^{(n)}$. This is called the Power Method.

The rate of convergence is controlled by $|\lambda_{n-1}|$. The closer to 1 this number is the slower the iterates converge. Also, if we know only that λ_n (for which $|\lambda_n| = \rho(A)$ is simple we can determine what it is by considering the *Rayleigh* quotient. Take

$$\rho_k = \frac{\langle A x^{(k)}, x^{(k)} \rangle}{\langle x^{(k)}, x^{(k)} \rangle}.$$

Then $\lim_{k \to \infty} \rho_k = \lambda_n$. Thus the multiple of $y^{(n)}$ is indeed λ_n.

To find intermediate eigenvalues and eigenvectors we apply an adaptation of the power method called the *orthogonalization method*. However, in order to adapt the power method to determine λ_{n-1}, our underlying assumption is that is also simple and morover $|\lambda_{n-2}| < |\lambda_{n-1}|$.

Assume $y^{(n)}$ and λ_n are known. Then we restart the iteration, taking the starting value

$$\hat{x}^{(0)} = x^{(0)} - \frac{\langle x^{(0)}, y^{(n)} \rangle}{\|y^{(n)}\|^2} y^{(n)}.$$

We know that the eigenvector $y^{(n-1)}$ pertaining to λ_{n-1} is orthogonal to $y^{(n)}$. Thus, in theory all of the iterates

$$\hat{x}^{(k+1)} = \frac{1}{\|\hat{x}^{(k)}\|} T \hat{x}^{(k)}$$

will remain orthogonal to $y^{(n)}$. Therefore,

$$\lim_{k \to \infty} \hat{x}^{(k)} = y^{(n-1)}.$$

—in theory. In practice, however, we must accept that $y^{(n)}$ has not been determined exactly. This means $\hat{x}^{(0)}$ has not been *purged* of all of $y^{(n)}$. By our previous reasoning, since λ_n is the dominant eigenvalue, the presence of $y^{(n)}$ will *creep* back into the iterates $\hat{x}^{(k)}$. To reduce the *contamination* it is best to *purify* the iterates $\hat{x}^{(k)}$ periodically by the reduction

$(\star)$
$$\hat{x}^{(k)} \longrightarrow \hat{x}^{(k)} - \frac{\langle \hat{x}^{(k)}, y^{(n)} \rangle}{\|y^{(n)}\|} y^{(n)}$$

before computing $\hat{x}^{(k+1)}$. The previous argument can be applied to prove that

$$(2) \quad \lim_{k \to \infty} x^{(k)} = y^{(n-1)}$$

$$(2) \quad \lim_{k \to \infty} \frac{\langle Ax^{(k)}, x^{(k)} \rangle}{\|x^{(k)}\|^2} = \lambda_{n-1}.$$

Additionally, even if we know $y^{(n)}$ exactly, round-off error would reinstate a $y^{(n)}$ component in our iterative computations. Thus the purification step above, $(\star)$, should be applied in *all* circumstances.

Finally, subsequent eigenvalues and eigenvectors may be determined by successive orthogonalizations. Again the eigenvalue simplicity and strict inequality is needed for convergence. Specifically, all eigenvectors can be determined if we assume that eigenvalues to be strictly ordered $|\lambda_1| < \cdots < |\lambda_{n-1}| < |\lambda_n| = \rho_n$. For example, we begin the iterations to determine $y^{(n-j)}$ with

$$\hat{x}^{(0)} = x^{(0)} - \sum_{i=0}^{j-1} \frac{\langle x^{(0)}, y^{(n-i)} \rangle}{\|y^{(n-i)}\|^2} y^{(n-i)}.$$

Don't forget the re-orthogonalizations periodically throughout the iterative process.

What can be done to find intermediate eigenvalues and eigenvectors in the case A is not symmetric? The method above fails, but a variation of it works.

What must be done is to generate the left and right eigenvectors, $w^{(n)}$ and $y^{(n)}$, for A. Use the same process. To compute $y^{(n-1)}$ and $w^{(n-1)}$ we

orthogonalize thusly:

$$\hat{x}^{(0)} = x^{(0)} - \frac{\langle x^{(0)}, w^{(n)} \rangle}{\|w^{(n)}\|^2} w^{(n)}$$

$$\hat{z}^{(0)} = z^{(0)} - \frac{\langle z^{(0)}, y^{(n)} \rangle}{\|y^{(n)}\|^2} y^{(n)}$$

where $z^{(0)}$ is the original starting value used to determine the left eigenvector $w^{(n)}$. Since we know that

$$\lim_{k \to \infty} x^{(k)} = \alpha y^{(n)}$$

it is easy to see that

$$\lim_{k \to \infty} Ax^{(k)} = \alpha \lambda_n y^{(n)}.$$

Therefore,

$$\lim_{k \to \infty} \frac{\langle Ax^{(k)}, x^{(k)} \rangle}{\langle x^{(k)}, x^{(k)} \rangle} = \lambda_n.$$

Example 5.2.1. Let T be the transformation of $R^2 \to R^2$ that rotates a vector by θ radians. Then it is clear that no matter what nonzero vector $x^{(0)}$ is selected the iterations $x^{(k+1)} = Tx^{(k)}$ will never converge. (Assume $\|x^{(0)}\| = 1$.) Now the matrix representation of T is

$$A_T = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad \text{rotates counterclockwise}$$

we have

$$p_{A_T}(\lambda) = \det \begin{bmatrix} \lambda - \cos\theta & \sin\theta \\ -\sin\theta & \lambda - \cos\theta \end{bmatrix} = (\lambda - \cos\theta)^2 + \sin^2\theta.$$

The spectrum of A_T is therefore

$$\lambda = \cos\theta \pm i\sin\theta.$$

Notice that the eigenvalues are discrete, but there are *two* eigenvalues with modulus $\rho(A_T) = 1$. The above results therefore do not apply.

Example 5.2.2. Although the previous example is not based on a symmetric matrix, it certainly illustrates non convergence of the power iterations. The even simpler Householder matrix $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ furnishes us with a symmetric matrix for which the iterations also do not converge. In this case, there are two eigenvalues with modulus equal to the spectral radius (± 1). With arbitrary starting vector $x^{(0)} = [a, \ b]^T$, it is obvious that the even iterations are $x^{(2i)} = [a, \ b]^T$ and the odd iterations are $x^{(2i-1)} = [a, \ -b]^T$.

Assuming again that the eigenvalues are distinct and even stronger, assuming that

$$|\lambda_1| < |\lambda_2| < \cdots < |\lambda_n|$$

we can apply the process above to extract all the eigenvalues (Rayleigh quotient) and the eigenvectors, one-by-one, when **A is symmetric**.

First of all, considering the matrix $A - \sigma I$ we can shift the eigenvalues to either the left or the right. Depending on the location of λ_n a sufficiently large $|\sigma|$ may be chosen so that $|\lambda_1 - \sigma| = \rho(A - \sigma I)$. The power method can be applied to determine $\lambda_1 - \sigma$ and hence λ_1.

5.3 Positive definite matrices

Of the many important subclasses of Hermitian matrices, there is one class that stands out.

Definition 5.3.1. We say that $A \in M_n(\mathbb{C})$ is *positive definite* if $\langle Ax, x \rangle > 0$ for every nonzero $x \in C_n$. Similarly, we say that $A \in M_n(C)$ is *positive semidefinite* if $\langle Ax, x \rangle \geq 0$ for every nonzero $x \in C_n$.

It is easy to see that for positive definite matrices all of the results are true

Theorem 5.3.1. *Let $A, B \in M_n(\mathbb{C})$. Then*

1. *If A is positive definite, then $\sigma(A) \subset R_n^+$*

2. *If A is positive definite, then A is invertible.*

3. *B^*B is positive semidefinite.*

4. *If B is invertible then B^*B is positive definite.*

5. *If $B \in M_n(C)$ is positive semidefinite, then $diag(B)$ is nonnegative, and $diag(B)$ is strictly positive when B is postive definite.*

The proofs are all routine. Of course, every diagonal matrix with non-negative entries is positive semidefinite.

Square roots

Given a real matrix $A \in M_n$. It is sometimes desired to determine a square root of A. By this we mean any matrix B for which $B^2 = A$. Moreover, if possible, it is desired that the square root be real. Our experience with numbers indicates that in order that a number have a positive square root, it must be positive. The analogue for matrices is the condition of being positive definite.

Theorem 5.3.2. *Let $A \in M_n$ be positive [semi-]definite. Then A has a real square root. Moreover, the square root can taken to be positive //[semi-]definite.*

Proof. We can write the diagonal matrix of the eigenvalues of A in the equation $A = P^{-1}DP$. Extract the positive square root of D as $D^{\frac{1}{2}} = \mathrm{diag}(\lambda_1^{1/2}, \ldots, \lambda_n^{1/2})$. Obviously $D^{\frac{1}{2}}D^{\frac{1}{2}} = D$. Now define $A^{\frac{1}{2}}$ by $A^{\frac{1}{2}} = P^{-1}D^{\frac{1}{2}}P$. This matrix is real. It is simple to check that $A^{\frac{1}{2}}A^{\frac{1}{2}} = A$, and that this particular square root is positive definite. $\square$

Clearly any real diagonalizable matrix with nonnegative eigenvectors has a real square root as well. However, beyond that conditions for determining existence let alone determination of square roots take us into a very specialized subject.

5.4   Singular Value Decomposition

Definition 5.4.1. For any $A \in M_{mn}$, the $n \times n$ Hermitian matrix A^*A is positive semi-definite. Denoting its eigenvalues by λ_j we called the values $\sqrt{\lambda_j}$ the **singular values** of A.

Because $r(A^*A) \leq \min(r(A^*), r(A)) \leq \min(m, n)$ there are at most $\min(m, n)$ nonzero singular values.

Lemma 5.4.1. *Let $A \in M_{mn}$. There is an orthonormal basis $\{u_1, \ldots, u_n\}$ of C_n such that $\{Au_1, \ldots, Au_n\}$ is orthogonal.*

Proof. Proof. Consider the $n \times n$ Hermitian matrix A^*A, and denote an orthonormal basis of its eigenvectors by $\{u_1, \ldots, u_n\}$. Then it is easy to see that $\{Au_1, \ldots, Au_n\}$ is an orthogonal set. For $\langle Au_j, Au_k \rangle = \langle A^*Au_j, u_k \rangle = \lambda_j \langle u_j, u_k \rangle = 0$. $\square$

Lemma 5.4.2. *Lemma 2 Let $A \in M_{mn}$ and an orthonormal basis $\{u_1, \ldots, u_n\}$ of C_n. Define*

$$v_j = \begin{cases} \frac{1}{\|Au_j\|}Au_j & if \quad \|Au_j\| \neq 0 \\ 0 & if \quad \|Au_j\| = 0 \end{cases}$$

Let $S = diag(\|Au_1\|, \ldots, \|Au_n\|)$, the $n \times n$ matrix U having rows given by the basis $\{u_1, \ldots, u_n\}$ and $\hat{V}$ the $m \times n$ matrix given by the columns $\{v_1, \ldots, v_n\}$. Then $A = \hat{V}SU$.

Proof. Proof. Consider

$$\hat{V}SUu_j = \begin{bmatrix} v_1 & v_2 & & v_n \\ \downarrow & \downarrow & & \downarrow \\ & & & \end{bmatrix} \begin{bmatrix} \|Au_1\| & 0 & \cdots & 0 \\ 0 & \|Au_2\| & 0 & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \|Au_n\| \end{bmatrix} \begin{bmatrix} u_1 & \longrightarrow \\ u_2 & \longrightarrow \\ \\ u_n & \longrightarrow \end{bmatrix} u_j$$

$$= \begin{bmatrix} v_1 & v_2 & & v_n \\ \downarrow & \downarrow & & \downarrow \\ & & & \end{bmatrix} \begin{bmatrix} \|Au_1\| & 0 & \cdots & 0 \\ 0 & \|Au_2\| & 0 & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \|Au_n\| \end{bmatrix} e_j$$

$$= \begin{bmatrix} v_1 & v_2 & & v_n \\ \downarrow & \downarrow & & \downarrow \\ & & & \end{bmatrix} \|Au_j\| e_j = \|Au_j\| v_j = Au_j$$

Thus both $\hat{V}SU$ and A have the same action on a basis. Therefore they are equal. $\square$

It is easy to see that $\|Au_j\| = \sqrt{\lambda_j}$, that is the singular values. While $A = \hat{V}SU$ could be called the singular value decomposition (SVD), what is usually offered at the SVD is small modification of it. Redefine the matrix U so that the first r columns pertain to the nonzero singularvalues. Define D to be the $m \times n$ matrix consisting of the non zero singular values in the

d_{jj} positions, and filled in with zeros else where. Define the matrix V to be the first r of the columns of $\hat{V}$ and if $r < m$ construct an additional $m - r$ orthonormal columns so that V is an orthonormal basis of $\mathbf{C}_n$. The resulting product VDU, called the **singular value decomposition** of A, is equal to A, and moreover it follows that V is $m \times m$, D is $m \times n$, and U is $n \times n$. This gives the following theorem

Theorem 5.4.1. *Let $A \in M_{mn}$. Then there is an $m \times m$ orthogonal matrix V, an $n \times n$ orthogonal matrix U, and an $m \times n$ matrix D with only diagonal entries such that $A = VDU$. The diagonal entries of D are the singular values of A and the rows of U are the eigenvectors of A^*A.*

Example 5.4.1. The singular value decomposition can be used for image compression. Here is the idea. Consider all the eigenvalues of A^*A and order them greatest to least. Zero the matrix S for all eigenvalues less than some threshold. Then in the reconstruction and transmission of the matrix, it is not necessary to include the vectors pertaining to these eigenvalues. In the example below, we have considered a 164×193 pixel image of C. F. Gauss (1777-1855) on the postage stamp issued by Germany on Feb. 23, 1955, to commemorate the centenary of death. Therefore its spectrum has 164 eigenvalues. The eigenvalues range from 26,603.0 to 1.895. A plot of the eigenvalues shown below. Now compress the image, retaining only a fraction of the eigenvalues by effectively zeroing the smaller eigenvalues.
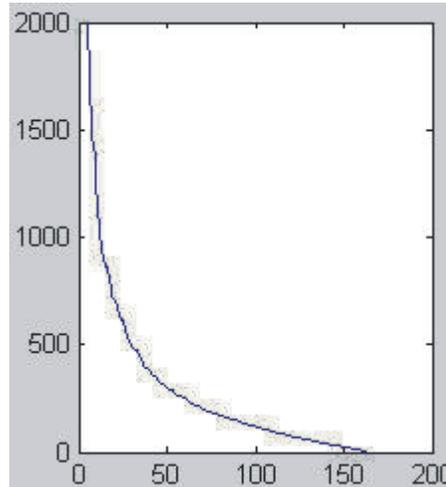
Note that the original image of the stamp has been enlarged and resampled for more accurate comparisons. This image (stored at 221 dpi) is displayed at effective 80 dpi with the enlargement. Below we show two plots where we have retained respectively 30% and 10% of the eigenvalues. There is an apparent drastic decline in the image quality at roughly 10:1 compression. In this image all eigenvalues smaller than $\lambda = 860$ have been zeroed.
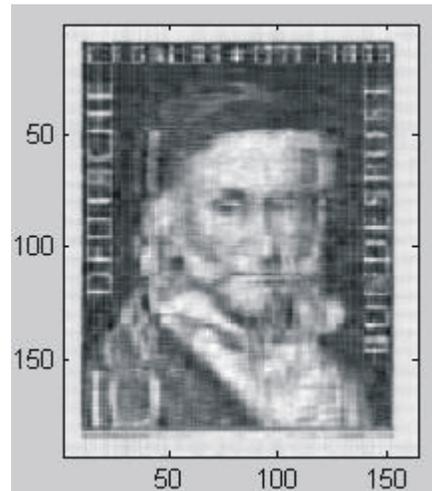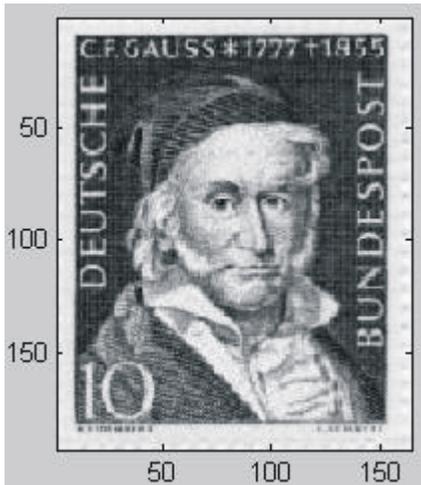
Using 48 of 164 eigenvalues   Using 16 of 164 eigenvalues

5.5   Exercises

1. Prove Theorem 5.3.1 (i).

2. Prove Theorem 5.3.1 (ii).

3. Suppose that $A, B \in M_n(\mathbb{C})$ are Hermitian. We will say $A \succcurlyeq 0$ if A is non-negative definite. Also, we say $A \succcurlyeq B$ if $A - B \succcurlyeq 0$. Is "$\succcurlyeq$" an equivalence relation? If $A \succcurlyeq B$ and $B \succcurlyeq C$ prove or disprove that $A \succcurlyeq C$.

4. Describe all Hermitian matrices of rank one.

5. Suppose that $A, B \in M_n(\mathbb{C})$ are Hermitian and positive definite. Find necessary and sufficient conditions for AB to be Hermitian and also positive definite.

6. For any matrix $A \in M_n(\mathbb{C})$ with eigenvalues $\lambda_i,\ i = 1, \ldots, n$. Prove that $\sum_{i=1}^{n} |\lambda_i|^2 = \sum_{i,j=1}^{n} |a_{ij}|^2$.