CHAPTER

1

BACKGROUND IN LINEAR ALGEBRA

This chapter gives an overview of the relevant concepts in linear algebra which are useful in later chapters. It begins with a review of basic matrix theory and introduces the elementary notation used throughout the book. The convergence analysis of iterative methods requires a good level of knowledge in mathematical analysis and in linear algebra. Traditionally, many of the concepts presented specifically for these analyses have been geared toward matrices arising from the discretization of Partial Differential Equations and basic relaxation-type methods. These concepts are now becoming less important because of the trend toward projection-type methods which have more robust convergence properties and require different analysis tools. The material covered in this chapter will be helpful in establishing some theory for the algorithms and defining the notation used throughout the book.

1.1 MATRICES

For the sake of generality, all vector spaces considered in this chapter are complex, unless otherwise stated. A complex $n \times m$ matrix A is an $n \times m$ array of complex numbers

$$a_{ij}, \;\; i = 1, \ldots, n, \; j = 1, \ldots, m.$$

The set of all $n \times m$ matrices is a complex vector space denoted by $\mathbb{C}^{n \times m}$. The main operations with matrices are the following:

- Addition: $C = A + B$, where A, B, and C are matrices of size $n \times m$ and

$$c_{ij} = a_{ij} + b_{ij}, \quad i = 1, 2, \ldots n, \quad j = 1, 2, \ldots m.$$

- Multiplication by a scalar: $C = \alpha A$, where

$$c_{ij} = \alpha \, a_{ij}, \quad i = 1, 2, \ldots n, \quad j = 1, 2, \ldots m.$$

- Multiplication by another matrix:

$$C = AB,$$

where $A \in \mathbb{C}^{n \times m}, B \in \mathbb{C}^{m \times p}, C \in \mathbb{C}^{n \times p}$, and

$$c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj}.$$

Sometimes, a notation with column vectors and row vectors is used. The column vector a_{*j} is the vector consisting of the j-th column of A,

$$a_{*j} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix}.$$

Similarly, the notation a_{i*} will denote the i-th row of the matrix A

$$a_{i*} = (a_{i1}, a_{i2}, \ldots, a_{im}).$$

For example, the following could be written

$$A = (a_{*1}, a_{*2}, \ldots, a_{*m}),$$

or

$$A = \begin{pmatrix} a_{1*} \\ a_{2*} \\ . \\ . \\ a_{n*} \end{pmatrix}.$$

The *transpose* of a matrix A in $\mathbb{C}^{n \times m}$ is a matrix C in $\mathbb{C}^{m \times n}$ whose elements are defined by $c_{ij} = a_{ji}, i = 1, \ldots, m, \ j = 1, \ldots, n$. It is denoted by A^T. It is often more relevant to use the *transpose conjugate* matrix denoted by A^H and defined by

$$A^H = \bar{A}^T = \overline{A^T},$$

in which the bar denotes the (element-wise) complex conjugation.

Matrices are strongly related to linear mappings between vector spaces of finite dimension. This is because they represent these mappings with respect to two given bases: one for the initial vector space and the other for the image vector space, or *range* of A.

1.2 SQUARE MATRICES AND EIGENVALUES

A matrix is *square* if it has the same number of columns and rows, i.e., if $m = n$. An important square matrix is the identity matrix

$$I = \{\delta_{ij}\}_{i,j=1,\ldots,n},$$

where δ_{ij} is the Kronecker symbol. The identity matrix satisfies the equality $AI = IA = A$ for every matrix A of size n. The inverse of a matrix, when it exists, is a matrix C such that

$$CA = AC = I.$$

The inverse of A is denoted by A^{-1}.

The *determinant* of a matrix may be defined in several ways. For simplicity, the following recursive definition is used here. The determinant of a 1×1 matrix (a) is defined as the scalar a. Then the determinant of an $n \times n$ matrix is given by

$$\det(A) = \sum_{j=1}^{n} (-1)^{j+1} a_{1j} \det(A_{1j}),$$

where A_{1j} is an $(n-1) \times (n-1)$ matrix obtained by deleting the first row and the j-th column of A. A matrix is said to be *singular* when $\det(A) = 0$ and *nonsingular* otherwise. We have the following simple properties:

- $\det(AB) = \det(BA)$.
- $\det(A^T) = \det(A)$.
- $\det(\alpha A) = \alpha^n \det(A)$.
- $\det(\bar{A}) = \overline{\det(A)}$.
- $\det(I) = 1$.

From the above definition of determinants it can be shown by induction that the function that maps a given complex value λ to the value $p_A(\lambda) = \det(A - \lambda I)$ is a polynomial of degree n; see Exercise 8. This is known as the *characteristic polynomial* of the matrix A.

DEFINITION 1.1 *A complex scalar λ is called an eigenvalue of the square matrix A if a nonzero vector u of $\mathbb{C}^n$ exists such that $Au = \lambda u$. The vector u is called an* eigenvector *of A associated with λ. The set of all the eigenvalues of A is called the spectrum of A and is denoted by $\sigma(A)$.*

A scalar λ is an eigenvalue of A if and only if $\det(A - \lambda I) \equiv p_A(\lambda) = 0$. That is true *if and only if* (*iff* thereafter) λ is a root of the characteristic polynomial. In particular, there are at most n distinct eigenvalues.

It is clear that a matrix is singular if and only if it admits zero as an eigenvalue. A well known result in linear algebra is stated in the following proposition.

PROPOSITION 1.1 *A matrix A is nonsingular if and only if it admits an inverse.*

Thus, the determinant of a matrix determines whether or not the matrix admits an inverse.

The maximum modulus of the eigenvalues is called *spectral radius* and is denoted by $\rho(A)$

$$\rho(A) = \max_{\lambda \in \sigma(A)} |\lambda|.$$

The *trace* of a matrix is equal to the sum of all its diagonal elements

$$\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}.$$

It can be easily shown that the trace of A is also equal to the sum of the eigenvalues of A counted with their multiplicities as roots of the characteristic polynomial.

PROPOSITION 1.2 *If λ is an eigenvalue of A, then $\bar{\lambda}$ is an eigenvalue of A^H. An eigenvector v of A^H associated with the eigenvalue $\bar{\lambda}$ is called a left eigenvector of A.*

When a distinction is necessary, an eigenvector of A is often called a right eigenvector. Therefore, the eigenvalue λ as well as the right and left eigenvectors, u and v, satisfy the relations

$$Au = \lambda u, \quad v^H A = \lambda v^H,$$

or, equivalently,

$$u^H A^H = \bar{\lambda} u^H, \quad A^H v = \bar{\lambda} v.$$

1.3 TYPES OF MATRICES

The choice of a method for solving linear systems will often depend on the structure of the matrix A. One of the most important properties of matrices is symmetry, because of its impact on the eigenstructure of A. A number of other classes of matrices also have particular eigenstructures. The most important ones are listed below:

- *Symmetric matrices:* $A^T = A$.
- *Hermitian matrices:* $A^H = A$.
- *Skew-symmetric matrices:* $A^T = -A$.
- *Skew-Hermitian matrices:* $A^H = -A$.
- *Normal matrices:* $A^H A = AA^H$.
- *Nonnegative matrices:* $a_{ij} \geq 0,\ i, j = 1, \ldots, n$ (similar definition for nonpositive, positive, and negative matrices).
- *Unitary matrices:* $Q^H Q = I$.

It is worth noting that a unitary matrix Q is a matrix whose inverse is its transpose conjugate Q^H, since

$$Q^H Q = I \quad \rightarrow \quad Q^{-1} = Q^H. \tag{1.1}$$

A matrix Q such that $Q^H Q$ is diagonal is often called orthogonal.

Some matrices have particular structures that are often convenient for computational purposes. The following list, though incomplete, gives an idea of these special matrices which play an important role in numerical analysis and scientific computing applications.

- *Diagonal matrices:* $a_{ij} = 0$ for $j \neq i$. Notation:

$$A = \text{diag}\ (a_{11}, a_{22}, \ldots, a_{nn}).$$

- *Upper triangular matrices:* $a_{ij} = 0$ for $i > j$.
- *Lower triangular matrices:* $a_{ij} = 0$ for $i < j$.
- *Upper bidiagonal matrices:* $a_{ij} = 0$ for $j \neq i$ or $j \neq i + 1$.
- *Lower bidiagonal matrices:* $a_{ij} = 0$ for $j \neq i$ or $j \neq i - 1$.
- *Tridiagonal matrices:* $a_{ij} = 0$ for any pair i, j such that $|j - i| > 1$. Notation:

$$A = \text{tridiag}\ (a_{i,i-1}, a_{ii}, a_{i,i+1}).$$

- *Banded matrices:* $a_{ij} \neq 0$ only if $i - m_l \leq j \leq i + m_u$, where m_l and m_u are two nonnegative integers. The number $m_l + m_u + 1$ is called the bandwidth of A.
- *Upper Hessenberg matrices:* $a_{ij} = 0$ for any pair i, j such that $i > j + 1$. Lower Hessenberg matrices can be defined similarly.
- *Outer product matrices:* $A = uv^H$, where both u and v are vectors.
- *Permutation matrices:* the columns of A are a permutation of the columns of the identity matrix.
- *Block diagonal matrices:* generalizes the diagonal matrix by replacing each diagonal entry by a matrix. Notation:

$$A = \text{diag}\ (A_{11}, A_{22}, \ldots, A_{nn}).$$

- *Block tridiagonal matrices:* generalizes the tridiagonal matrix by replacing each nonzero entry by a square matrix. Notation:

$$A = \text{tridiag}\ (A_{i,i-1}, A_{ii}, A_{i,i+1}).$$

The above properties emphasize structure, i.e., positions of the nonzero elements with respect to the zeros. Also, they assume that there are many zero elements or that the matrix is of low rank. This is in contrast with the classifications listed earlier, such as symmetry or normality.

1.4 VECTOR INNER PRODUCTS AND NORMS

An inner product on a (complex) vector space $\mathbb{X}$ is any mapping s from $\mathbb{X} \times \mathbb{X}$ into $\mathbb{C}$,

$$x \in \mathbb{X}, y \in \mathbb{X} \quad \rightarrow \quad s(x, y) \in \mathbb{C},$$

which satisfies the following conditions:

1. $s(x, y)$ is linear with respect to x, i.e.,

$$s(\lambda_1 x_1 + \lambda_2 x_2, y) = \lambda_1 s(x_1, y) + \lambda_2 s(x_2, y), \quad \forall\, x_1, x_2 \in \mathbb{X}, \forall\, \lambda_1, \lambda_2 \in \mathbb{C}.$$

2. $s(x, y)$ is *Hermitian*, i.e.,

$$s(y, x) = \overline{s(x, y)}, \quad \forall\, x, y \in \mathbb{X}.$$

3. $s(x, y)$ is *positive definite*, i.e.,

$$s(x, x) > 0, \quad \forall\, x \neq 0.$$

Note that (2) implies that $s(x, x)$ is real and therefore, (3) adds the constraint that $s(x, x)$ must also be positive for any nonzero x. For any x and y,

$$s(x, 0) = s(x, 0.y) = 0.s(x, y) = 0.$$

Similarly, $s(0, y) = 0$ for any y. Hence, $s(0, y) = s(x, 0) = 0$ for any x and y. In particular the condition (3) can be rewritten as

$$s(x, x) \geq 0 \quad \text{and} \quad s(x, x) = 0 \quad \text{iff} \quad x = 0,$$

as can be readily shown. A useful relation satisfied by any inner product is the so-called Cauchy-Schwartz inequality:

$$|s(x, y)|^2 \leq s(x, x)\; s(y, y). \tag{1.2}$$

The proof of this inequality begins by expanding $s(x - \lambda y, x - \lambda y)$ using the properties of s,

$$s(x - \lambda y, x - \lambda y) = s(x, x) - \bar{\lambda} s(x, y) - \lambda s(y, x) + |\lambda|^2 s(y, y).$$

If $y = 0$ then the inequality is trivially satisfied. Assume that $y \neq 0$ and take $\lambda = s(x, y)/s(y, y)$. Then $s(x - \lambda y, x - \lambda y) \geq 0$ shows the above equality

$$\begin{aligned} 0 \leq s(x - \lambda y, x - \lambda y) &= s(x, x) - 2\frac{|s(x, y)|^2}{s(y, y)} + \frac{|s(x, y)|^2}{s(y, y)} \\ &= s(x, x) - \frac{|s(x, y)|^2}{s(y, y)}, \end{aligned}$$

which yields the result (1.2).

In the particular case of the vector space $\mathbb{X} = \mathbb{C}^n$, a "canonical" inner product is the *Euclidean inner product*. The Euclidean inner product of two vectors $x = (x_i)_{i=1,\ldots,n}$ and

$y = (y_i)_{i=1,\ldots,n}$ of $\mathbb{C}^n$ is defined by

$$(x, y) = \sum_{i=1}^{n} x_i \bar{y}_i, \tag{1.3}$$

which is often rewritten in matrix notation as

$$(x, y) = y^H x. \tag{1.4}$$

It is easy to verify that this mapping does indeed satisfy the three conditions required for inner products, listed above. A fundamental property of the Euclidean inner product in matrix computations is the simple relation

$$(Ax, y) = (x, A^H y), \quad \forall\, x, y \in \mathbb{C}^n. \tag{1.5}$$

The proof of this is straightforward. The *adjoint* of A *with respect to an arbitrary inner product* is a matrix B such that $(Ax, y) = (x, By)$ for all pairs of vectors x and y. A matrix is *self-adjoint*, or Hermitian with respect to this inner product, if it is equal to its adjoint.

The following proposition is a consequence of the equality (1.5).

PROPOSITION 1.3 *Unitary matrices preserve the Euclidean inner product, i.e.,*

$$(Qx, Qy) = (x, y)$$

for any unitary matrix Q and any vectors x and y.

Proof. Indeed, $(Qx, Qy) = (x, Q^H Qy) = (x, y)$. ■

A vector norm on a vector space $\mathbb{X}$ is a real-valued function $x \to \|x\|$ on $\mathbb{X}$, which satisfies the following three conditions:

1. $\|x\| \geq 0, \quad \forall\, x \in \mathbb{X}, \quad$ and $\quad \|x\| = 0$ iff $x = 0$.
2. $\|\alpha x\| = |\alpha| \|x\|, \quad \forall\, x \in \mathbb{X}, \quad \forall \alpha \in \mathbb{C}$.
3. $\|x + y\| \leq \|x\| + \|y\|, \quad \forall\, x, y \in \mathbb{X}$.

For the particular case when $\mathbb{X} = \mathbb{C}^n$, we can associate with the inner product (1.3) the *Euclidean norm* of a complex vector defined by

$$\|x\|_2 = (x, x)^{1/2}.$$

It follows from Proposition 1.3 that a unitary matrix preserves the Euclidean norm metric, i.e.,

$$\|Qx\|_2 = \|x\|_2, \ \forall\, x.$$

The linear transformation associated with a unitary matrix Q is therefore an *isometry*.

The most commonly used vector norms in numerical linear algebra are special cases of the Hölder norms

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}. \tag{1.6}$$

Note that the limit of $\|x\|_p$ when p tends to infinity exists and is equal to the maximum modulus of the x_i's. This defines a norm denoted by $\|.\|_\infty$. The cases $p = 1$, $p = 2$, and $p = \infty$ lead to the most important norms in practice,

$$\|x\|_1 = |x_1| + |x_2| + \cdots + |x_n|,$$
$$\|x\|_2 = \left[|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2\right]^{1/2},$$
$$\|x\|_\infty = \max_{i=1,\ldots,n} |x_i|.$$

The Cauchy-Schwartz inequality of (1.2) becomes

$$|(x, y)| \le \|x\|_2 \|y\|_2.$$

1.5 MATRIX NORMS

For a general matrix A in $\mathbb{C}^{n\times m}$, we define the following special set of norms

$$\|A\|_{pq} = \max_{x\in\mathbb{C}^m,\ x\neq 0} \frac{\|Ax\|_p}{\|x\|_q}. \tag{1.7}$$

The norm $\|.\|_{pq}$ is *induced* by the two norms $\|.\|_p$ and $\|.\|_q$. These norms satisfy the usual properties of norms, i.e.,

$$\|A\| \ge 0, \quad \forall\, A \in \mathbb{C}^{n\times m}, \quad \text{and} \quad \|A\| = 0 \quad \text{iff} \quad A = 0$$
$$\|\alpha A\| = |\alpha| \|A\|, \forall\, A \in \mathbb{C}^{n\times m}, \quad \forall\, \alpha \in \mathbb{C}$$
$$\|A + B\| \le \|A\| + \|B\|, \quad \forall\, A, B \in \mathbb{C}^{n\times m}.$$

The most important cases are again those associated with $p, q = 1, 2, \infty$. The case $q = p$ is of particular interest and the associated norm $\|.\|_{pq}$ is simply denoted by $\|.\|_p$ and called a "p-norm." A fundamental property of a p-norm is that

$$\|AB\|_p \le \|A\|_p \|B\|_p,$$

an immediate consequence of the definition (1.7). Matrix norms that satisfy the above property are sometimes called *consistent*. A result of consistency is that for any square matrix A,

$$\|A^k\|_p \le \|A\|_p^k.$$

In particular the matrix A^k converges to zero if *any* of its p-norms is less than 1.

The Frobenius norm of a matrix is defined by

$$\|A\|_F = \left(\sum_{j=1}^{m} \sum_{i=1}^{n} |a_{ij}|^2\right)^{1/2}. \tag{1.8}$$

This can be viewed as the 2-norm of the column (or row) vector in $\mathbb{C}^{n^2}$ consisting of all the columns (respectively rows) of A listed from 1 to m (respectively 1 to n.) It can be shown that this norm is also consistent, in spite of the fact that it is not induced by a pair of vector norms, i.e., it is not derived from a formula of the form (1.7); see Exercise 5. However, it does not satisfy some of the other properties of the p-norms. For example, the Frobenius norm of the identity matrix is not equal to one. To avoid these difficulties, *we will only use the term matrix norm for a norm that is induced by two norms as in the definition (1.7).* Thus, we will not consider the Frobenius norm to be a proper matrix norm, according to our conventions, even though it is consistent.

The following equalities satisfied by the matrix norms defined above lead to alternative definitions that are often easier to work with:

$$\|A\|_1 = \max_{j=1,\ldots,m} \sum_{i=1}^{n} |a_{ij}|, \tag{1.9}$$

$$\|A\|_\infty = \max_{i=1,\ldots,n} \sum_{j=1}^{m} |a_{ij}|, \tag{1.10}$$

$$\|A\|_2 = \left[\rho(A^H A)\right]^{1/2} = \left[\rho(AA^H)\right]^{1/2}, \tag{1.11}$$

$$\|A\|_F = \left[\mathrm{tr}(A^H A)\right]^{1/2} = \left[\mathrm{tr}(AA^H)\right]^{1/2}. \tag{1.12}$$

As will be shown later, the eigenvalues of $A^H A$ are nonnegative. Their square roots are called *singular values* of A and are denoted by $\sigma_i, i = 1, \ldots, m$. Thus, the relation (1.11) states that $\|A\|_2$ is equal to σ_1, the largest singular value of A.

Example 1.1 From the relation (1.11), it is clear that the spectral radius $\rho(A)$ is equal to the 2-norm of a matrix when the matrix is Hermitian. However, it is not a matrix norm in general. For example, the first property of norms is not satisfied, since for

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

we have $\rho(A) = 0$ while $A \neq 0$. Also, the triangle inequality is not satisfied for the pair A, and $B = A^T$ where A is defined above. Indeed,

$$\rho(A + B) = 1 \quad \text{while} \quad \rho(A) + \rho(B) = 0.$$

1.6 SUBSPACES, RANGE, AND KERNEL

A subspace of $\mathbb{C}^n$ is a subset of $\mathbb{C}^n$ that is also a complex vector space. The set of all linear combinations of a set of vectors $G = \{a_1, a_2, \ldots, a_q\}$ of $\mathbb{C}^n$ is a vector subspace called the linear span of G,

$$\mathrm{span}\{G\} = \mathrm{span}\{a_1, a_2, \ldots, a_q\}$$

$$= \left\{ z \in \mathbb{C}^n \;\middle|\; z = \sum_{i=1}^{q} \alpha_i a_i;\; \{\alpha_i\}_{i=1,\ldots,q} \in \mathbb{C}^q \right\}.$$

If the a_i's are linearly independent, then each vector of $\mathrm{span}\{G\}$ admits a unique expression as a linear combination of the a_i's. The set G is then called a *basis* of the subspace $\mathrm{span}\{G\}$.

Given two vector subspaces S_1 and S_2, their *sum* S is a subspace defined as the set of all vectors that are equal to the sum of a vector of S_1 and a vector of S_2. The intersection of two subspaces is also a subspace. If the intersection of S_1 and S_2 is reduced to $\{0\}$, then the sum of S_1 and S_2 is called their direct sum and is denoted by $S = S_1 \bigoplus S_2$. When S is equal to $\mathbb{C}^n$, then every vector x of $\mathbb{C}^n$ can be written in a unique way as the sum of an element x_1 of S_1 and an element x_2 of S_2. The transformation P that maps x into x_1 is a linear transformation that is *idempotent*, i.e., such that $P^2 = P$. It is called a *projector* onto S_1 along S_2.

Two important subspaces that are associated with a matrix A of $\mathbb{C}^{n\times m}$ are its *range*, defined by

$$\mathrm{Ran}(A) = \{Ax \mid x \in \mathbb{C}^m\}, \tag{1.13}$$

and its *kernel* or *null space*

$$\mathrm{Ker}(A) = \{x \in \mathbb{C}^m \mid Ax = 0\}.$$

The range of A is clearly equal to the linear *span* of its columns. The *rank* of a matrix is equal to the dimension of the range of A, i.e., to the number of linearly independent columns. This *column rank* is equal to the *row rank*, the number of linearly independent rows of A. A matrix in $\mathbb{C}^{n\times m}$ is of *full rank* when its rank is equal to the smallest of m and n.

A subspace S is said to be *invariant* under a (square) matrix A whenever $AS \subset S$. In particular for any eigenvalue λ of A the subspace $\mathrm{Ker}(A - \lambda I)$ is invariant under A. The subspace $\mathrm{Ker}(A - \lambda I)$ is called the eigenspace associated with λ and consists of all the eigenvectors of A associated with λ, in addition to the zero-vector.

1.7 ORTHOGONAL VECTORS AND SUBSPACES

A set of vectors $G = \{a_1, a_2, \ldots, a_r\}$ is said to be *orthogonal* if

$$(a_i, a_j) = 0 \quad \text{when} \quad i \neq j.$$

It is *orthonormal* if, in addition, every vector of G has a 2-norm equal to unity. A vector that is orthogonal to all the vectors of a subspace S is said to be orthogonal to this subspace. The set of all the vectors that are orthogonal to S is a vector subspace called the *orthogonal complement* of S and denoted by $S^\perp$. The space $\mathbb{C}^n$ is the direct sum of S and its orthogonal complement. Thus, any vector x can be written in a unique fashion as the sum of a vector in S and a vector in $S^\perp$. The operator which maps x into its component in the subspace S is the *orthogonal projector* onto S.

Every subspace admits an orthonormal basis which is obtained by taking any basis and "orthonormalizing" it. The orthonormalization can be achieved by an algorithm known as the Gram-Schmidt process which we now describe. Given a set of linearly independent vectors $\{x_1, x_2, \ldots, x_r\}$, first normalize the vector x_1, which means divide it by its 2-norm, to obtain the scaled vector q_1 of norm unity. Then x_2 is orthogonalized against the vector q_1 by subtracting from x_2 a multiple of q_1 to make the resulting vector orthogonal to q_1, i.e.,

$$x_2 \leftarrow x_2 - (x_2, q_1) q_1 .$$

The resulting vector is again normalized to yield the second vector q_2. The i-th step of the Gram-Schmidt process consists of orthogonalizing the vector x_i against all previous vectors q_j.

ALGORITHM 1.1: Gram-Schmidt

1. *Compute* $r_{11} := \|x_1\|_2$. *If* $r_{11} = 0$ *Stop, else compute* $q_1 := x_1/r_{11}$.
2. *For* $j = 2, \ldots, r$ *Do:*
3. *Compute* $r_{ij} := (x_j, q_i)$ *, for* $i = 1, 2, \ldots, j-1$
4. $\hat{q} := x_j - \sum_{i=1}^{j-1} r_{ij} q_i$
5. $r_{jj} := \|\hat{q}\|_2$,
6. *If* $r_{jj} = 0$ *then Stop, else* $q_j := \hat{q}/r_{jj}$
7. *EndDo*

It is easy to prove that the above algorithm will not break down, i.e., all r steps will be completed if and only if the set of vectors $x_1, x_2, \ldots, x_r$ is linearly independent. From lines 4 and 5, it is clear that at every step of the algorithm the following relation holds:

$$x_j = \sum_{i=1}^{j} r_{ij} q_i .$$

If $X = [x_1, x_2, \ldots, x_r]$, $Q = [q_1, q_2, \ldots, q_r]$, and if R denotes the $r \times r$ upper triangular matrix whose nonzero elements are the r_{ij} defined in the algorithm, then the above relation can be written as

$$X = QR. \tag{1.14}$$

This is called the QR decomposition of the $n \times r$ matrix X. From what was said above, the QR decomposition of a matrix exists whenever the column vectors of X form a linearly independent set of vectors.

The above algorithm is the standard Gram-Schmidt process. There are alternative formulations of the algorithm which have better numerical properties. The best known of these is the Modified Gram-Schmidt (MGS) algorithm.

ALGORITHM 1.2: Modified Gram-Schmidt

1. *Define* $r_{11} := \|x_1\|_2$. *If* $r_{11} = 0$ *Stop, else* $q_1 := x_1/r_{11}$.
2. *For* $j = 2, \ldots, r$ *Do:*

3. *Define* $\hat{q} := x_j$
4. *For* $i = 1, \ldots, j-1$, *Do:*
5. $r_{ij} := (\hat{q}, q_i)$
6. $\hat{q} := \hat{q} - r_{ij} q_i$
7. *EndDo*
8. *Compute* $r_{jj} := \|\hat{q}\|_2$,
9. *If* $r_{jj} = 0$ *then Stop, else* $q_j := \hat{q}/r_{jj}$
10. *EndDo*

Yet another alternative for orthogonalizing a sequence of vectors is the Householder algorithm. This technique uses Householder *reflectors*, i.e., matrices of the form

$$P = I - 2ww^T, \tag{1.15}$$

in which w is a vector of 2-norm unity. Geometrically, the vector Px represents a mirror image of x with respect to the hyperplane $\mathrm{span}\{w\}^\perp$.

To describe the Householder orthogonalization process, the problem can be formulated as that of finding a QR factorization of a given $n \times m$ matrix X. For any vector x, the vector w for the Householder transformation (1.15) is selected in such a way that

$$Px = \alpha e_1,$$

where α is a scalar. Writing $(I - 2ww^T)x = \alpha e_1$ yields

$$2w^T x \, w = x - \alpha e_1. \tag{1.16}$$

This shows that the desired w is a multiple of the vector $x - \alpha e_1$,

$$w = \pm \frac{x - \alpha e_1}{\|x - \alpha e_1\|_2}.$$

For (1.16) to be satisfied, we must impose the condition

$$2(x - \alpha e_1)^T x = \|x - \alpha e_1\|_2^2$$

which gives $2(\|x\|_1^2 - \alpha \xi_1) = \|x\|_2^2 - 2\alpha\xi_1 + \alpha^2$, where $\xi_1 \equiv e_1^T x$ is the first component of the vector x. Therefore, it is necessary that

$$\alpha = \pm \|x\|_2.$$

In order to avoid that the resulting vector w be small, it is customary to take

$$\alpha = -\mathrm{sign}(\xi_1)\|x\|_2,$$

which yields

$$w = \frac{x + \mathrm{sign}(\xi_1)\|x\|_2 e_1}{\|x + \mathrm{sign}(\xi_1)\|x\|_2 e_1\|_2}. \tag{1.17}$$

Given an $n \times m$ matrix, its first column can be transformed to a multiple of the column e_1, by premultiplying it by a Householder matrix P_1,

$$X_1 \equiv P_1 X, \qquad X_1 e_1 = \alpha e_1.$$

Assume, inductively, that the matrix X has been transformed in $k-1$ successive steps into the partially upper triangular form

$$X_k \equiv P_{k-1} \dots P_1 X_1 = \begin{pmatrix} x_{11} & x_{12} & x_{13} & \cdots & \cdots & \cdots & x_{1m} \\ & x_{22} & x_{23} & \cdots & \cdots & \cdots & x_{2m} \\ & & x_{33} & \cdots & \cdots & \cdots & x_{3m} \\ & & & \ddots & \cdots & \cdots & \vdots \\ & & & & x_{kk} & \cdots & \vdots \\ & & & & x_{k+1,k} & \cdots & x_{k+1,m} \\ & & & & \vdots & \vdots & \vdots \\ & & & & x_{n,k} & \cdots & x_{n,m} \end{pmatrix}.$$

This matrix is upper triangular up to column number $k-1$. To advance by one step, it must be transformed into one which is upper triangular up the k-th column, leaving the previous columns in the same form. To leave the first $k-1$ columns unchanged, select a w vector which has zeros in positions 1 through $k-1$. So the next Householder reflector matrix is defined as

$$P_k = I - 2w_k w_k^T, \tag{1.18}$$

in which the vector w_k is defined as

$$w_k = \frac{z}{\|z\|_2}, \tag{1.19}$$

where the components of the vector z are given by

$$z_i = \begin{cases} 0 & \text{if} \quad i < k \\ \beta + x_{ii} & \text{if} \quad i = k \\ x_{ik} & \text{if} \quad i > k \end{cases} \tag{1.20}$$

with

$$\beta = \operatorname{sign}(x_{kk}) \times \left(\sum_{i=k}^{n} x_{ik}^2 \right)^{1/2}. \tag{1.21}$$

We note in passing that the premultiplication of a matrix X by a Householder transform requires only a rank-one update since,

$$(I - 2ww^T)X = X - wv^T \quad \text{where} \quad v = 2X^T w.$$

Therefore, the Householder matrices need not, and should not, be explicitly formed. In addition, the vectors w need not be explicitly scaled.

Assume now that $m-1$ Householder transforms have been applied to a certain matrix

X of dimension $n \times m$, to reduce it into the upper triangular form,

$$X_m \equiv P_{m-1}P_{m-2}\dots P_1 X = \begin{pmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1m} \\ & x_{22} & x_{23} & \cdots & x_{2m} \\ & & x_{33} & \cdots & x_{3m} \\ & & & \ddots & \vdots \\ & & & & x_{m,m} \\ & & & & 0 \\ & & & & \vdots \\ & & & & \vdots \end{pmatrix}. \tag{1.22}$$

Recall that our initial goal was to obtain a QR factorization of X. We now wish to recover the Q and R matrices from the P_k's and the above matrix. If we denote by P the product of the P_i on the left-side of (1.22), then (1.22) becomes

$$PX = \begin{pmatrix} R \\ O \end{pmatrix}, \tag{1.23}$$

in which R is an $m \times m$ upper triangular matrix, and O is an $(n-m) \times m$ zero block. Since P is unitary, its inverse is equal to its transpose and, as a result,

$$X = P^T \begin{pmatrix} R \\ O \end{pmatrix} = P_1 P_2 \dots P_{m-1} \begin{pmatrix} R \\ O \end{pmatrix}.$$

If E_m is the matrix of size $n \times m$ which consists of the first m columns of the identity matrix, then the above equality translates into

$$X = P^T E_m R.$$

The matrix $Q = P^T E_m$ represents the m first columns of P^T. Since

$$Q^T Q = E_m^T P P^T E_m = I,$$

Q and R are the matrices sought. In summary,

$$X = QR,$$

in which R is the triangular matrix obtained from the Householder reduction of X (see (1.22) and (1.23)) and

$$Qe_j = P_1 P_2 \dots P_{m-1} e_j.$$

ALGORITHM 1.3: Householder Orthogonalization

1. *Define* $X = [x_1, \dots, x_m]$
2. *For* $k = 1, \dots, m$ *Do:*
3. *If* $k > 1$ *compute* $r_k := P_{k-1}P_{k-2}\dots P_1 x_k$
4. *Compute* w_k *using (1.19), (1.20), (1.21)*
5. *Compute* $r_k := P_k r_k$ *with* $P_k = I - 2w_k w_k^T$
6. *Compute* $q_k = P_1 P_2 \dots P_k e_k$
7. *EndDo*

Note that line 6 can be omitted since the q_i are not needed in the execution of the next steps. It must be executed only when the matrix Q is needed at the completion of the algorithm. Also, the operation in line 5 consists only of zeroing the components $k+1, \ldots, n$ and updating the k-th component of r_k. In practice, a work vector can be used for r_k and its nonzero components after this step can be saved into an upper triangular matrix. Since the components 1 through k of the vector w_k are zero, the upper triangular matrix R can be saved in those zero locations which would otherwise be unused.

1.8 CANONICAL FORMS OF MATRICES

This section discusses the reduction of square matrices into matrices that have simpler forms, such as diagonal, bidiagonal, or triangular. Reduction means a transformation that preserves the eigenvalues of a matrix.

DEFINITION 1.2 *Two matrices A and B are said to be similar if there is a nonsingular matrix X such that*

$$A = XBX^{-1}.$$

The mapping $B \to A$ is called a similarity transformation.

It is clear that *similarity* is an equivalence relation. Similarity transformations preserve the eigenvalues of matrices. An eigenvector u_B of B is transformed into the eigenvector $u_A = Xu_B$ of A. In effect, a similarity transformation amounts to representing the matrix B in a different basis.

We now introduce some terminology.

1. An eigenvalue λ of A has *algebraic multiplicity* μ, if it is a root of multiplicity μ of the characteristic polynomial.
2. If an eigenvalue is of algebraic multiplicity one, it is said to be *simple*. A nonsimple eigenvalue is *multiple*.
3. The *geometric multiplicity* γ of an eigenvalue λ of A is the maximum number of independent eigenvectors associated with it. In other words, the geometric multiplicity γ is the dimension of the eigenspace $\mathrm{Ker}\,(A - \lambda I)$.
4. A matrix is *derogatory* if the geometric multiplicity of at least one of its eigenvalues is larger than one.
5. An eigenvalue is *semisimple* if its algebraic multiplicity is equal to its geometric multiplicity. An eigenvalue that is not semisimple is called *defective*.

Often, $\lambda_1, \lambda_2, \ldots, \lambda_p$ ($p \leq n$) are used to denote the *distinct* eigenvalues of A. It is easy to show that the characteristic polynomials of two similar matrices are identical; see Exercise 9. Therefore, the eigenvalues of two similar matrices are equal and so are their algebraic multiplicities. Moreover, if v is an eigenvector of B, then Xv is an eigenvector of A and, conversely, if y is an eigenvector of A then $X^{-1}y$ is an eigenvector of B. As a result the number of independent eigenvectors associated with a given eigenvalue is the same for two similar matrices, i.e., their geometric multiplicity is also the same.

1.8.1 REDUCTION TO THE DIAGONAL FORM

The simplest form in which a matrix can be reduced is undoubtedly the diagonal form. Unfortunately, this reduction is not always possible. A matrix that can be reduced to the diagonal form is called *diagonalizable*. The following theorem characterizes such matrices.

THEOREM 1.1 *A matrix of dimension n is diagonalizable if and only if it has n linearly independent eigenvectors.*

Proof. A matrix A is diagonalizable if and only if there exists a nonsingular matrix X and a diagonal matrix D such that $A = XDX^{-1}$, or equivalently $AX = XD$, where D is a diagonal matrix. This is equivalent to saying that n linearly independent vectors exist — the n column-vectors of X — such that $Ax_i = d_i x_i$. Each of these column-vectors is an eigenvector of A. ■

A matrix that is diagonalizable has only semisimple eigenvalues. Conversely, if all the eigenvalues of a matrix A are semisimple, then A has n eigenvectors. It can be easily shown that these eigenvectors are linearly independent; see Exercise 2. As a result, we have the following proposition.

PROPOSITION 1.4 *A matrix is diagonalizable if and only if all its eigenvalues are semisimple.*

Since every simple eigenvalue is semisimple, an immediate corollary of the above result is: When A has n distinct eigenvalues, then it is diagonalizable.

1.8.2 THE JORDAN CANONICAL FORM

From the theoretical viewpoint, one of the most important canonical forms of matrices is the well known Jordan form. A full development of the steps leading to the Jordan form is beyond the scope of this book. Only the main theorem is stated. Details, including the proof, can be found in standard books of linear algebra such as [117]. In the following, m_i refers to the algebraic multiplicity of the individual eigenvalue λ_i and l_i is the *index* of the eigenvalue, i.e., the smallest integer for which $\mathrm{Ker}(A - \lambda_i I)^{l_i+1} = \mathrm{Ker}(A - \lambda_i I)^{l_i}$.

THEOREM 1.2 *Any matrix A can be reduced to a block diagonal matrix consisting of p diagonal blocks, each associated with a distinct eigenvalue λ_i. Each of these diagonal blocks has itself a block diagonal structure consisting of γ_i sub-blocks, where γ_i is the geometric multiplicity of the eigenvalue λ_i. Each of the sub-blocks, referred to as a Jordan* *block, is an upper bidiagonal matrix of size not exceeding $l_i \le m_i$, with the constant λ_i on the diagonal and the constant one on the super diagonal.*

The i-th diagonal block, $i = 1, \ldots, p$, is known as the i-th Jordan submatrix (sometimes "Jordan Box"). The Jordan submatrix number i starts in column $j_i \equiv m_1 + m_2 + \cdots + m_{i-1} + 1$. Thus,

$$X^{-1}AX = J = \begin{pmatrix} J_1 & & & & & \\ & J_2 & & & & \\ & & \ddots & & & \\ & & & J_i & & \\ & & & & \ddots & \\ & & & & & J_p \end{pmatrix},$$

where each J_i is associated with λ_i and is of size m_i the algebraic multiplicity of λ_i. It has itself the following structure,

$$J_i = \begin{pmatrix} J_{i1} & & & \\ & J_{i2} & & \\ & & \ddots & \\ & & & J_{i\gamma_i} \end{pmatrix} \text{ with } J_{ik} = \begin{pmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_i & 1 \\ & & & \lambda_i \end{pmatrix}.$$

Each of the blocks J_{ik} corresponds to a different eigenvector associated with the eigenvalue λ_i. Its size l_i is the index of λ_i.

1.8.3 THE SCHUR CANONICAL FORM

Here, it will be shown that any matrix is unitarily similar to an upper triangular matrix. The only result needed to prove the following theorem is that any vector of 2-norm one can be completed by $n-1$ additional vectors to form an orthonormal basis of $\mathbb{C}^n$.

THEOREM 1.3 *For any square matrix A, there exists a unitary matrix Q such that*

$$Q^H AQ = R$$

is upper triangular.

Proof. The proof is by induction over the dimension n. The result is trivial for $n = 1$. Assume that it is true for $n-1$ and consider any matrix A of size n. The matrix admits at least one eigenvector u that is associated with an eigenvalue λ. Also assume without loss of generality that $\|u\|_2 = 1$. First, complete the vector u into an orthonormal set, i.e., find an $n \times (n-1)$ matrix V such that the $n \times n$ matrix $U = [u, V]$ is unitary. Then $AU = [\lambda u, AV]$ and hence,

$$U^H AU = \begin{bmatrix} u^H \\ V^H \end{bmatrix} [\lambda u, AV] = \begin{pmatrix} \lambda & u^H AV \\ 0 & V^H AV \end{pmatrix}. \tag{1.24}$$

Now use the induction hypothesis for the $(n-1) \times (n-1)$ matrix $B = V^H AV$: There exists an $(n-1) \times (n-1)$ unitary matrix Q_1 such that $Q_1^H BQ_1 = R_1$ is upper triangular.

Define the $n \times n$ matrix

$$\hat{Q}_1 = \begin{pmatrix} 1 & 0 \\ 0 & Q_1 \end{pmatrix}$$

and multiply both members of (1.24) by $\hat{Q}_1^H$ from the left and $\hat{Q}_1$ from the right. The resulting matrix is clearly upper triangular and this shows that the result is true for A, with $Q = \hat{Q}_1 U$ which is a unitary $n \times n$ matrix. ∎

A simpler proof that uses the Jordan canonical form and the QR decomposition is the subject of Exercise 7. Since the matrix R is triangular and similar to A, its diagonal elements are equal to the eigenvalues of A ordered in a certain manner. In fact, it is easy to extend the proof of the theorem to show that this factorization can be obtained with *any order* for the eigenvalues. Despite its simplicity, the above theorem has far-reaching consequences, some of which will be examined in the next section.

It is important to note that for any $k \leq n$, the subspace spanned by the first k columns of Q is invariant under A. Indeed, the relation $AQ = QR$ implies that for $1 \leq j \leq k$, we have

$$Aq_j = \sum_{i=1}^{i=j} r_{ij} q_i .$$

If we let $Q_k = [q_1, q_2, \ldots, q_k]$ and if R_k is the principal leading submatrix of dimension k of R, the above relation can be rewritten as

$$AQ_k = Q_k R_k,$$

which is known as the partial Schur decomposition of A. The simplest case of this decomposition is when $k = 1$, in which case q_1 is an eigenvector. The vectors q_i are usually called Schur vectors. Schur vectors are not unique and depend, in particular, on the order chosen for the eigenvalues.

A slight variation on the Schur canonical form is the quasi-Schur form, also called the real Schur form. Here, diagonal blocks of size 2×2 are allowed in the upper triangular matrix R. The reason for this is to avoid complex arithmetic when the original matrix is real. A 2×2 block is associated with each complex conjugate pair of eigenvalues of the matrix.

Example 1.2 Consider the 3×3 matrix

$$A = \begin{pmatrix} 1 & 10 & 0 \\ -1 & 3 & 1 \\ -1 & 0 & 1 \end{pmatrix}.$$

The matrix A has the pair of complex conjugate eigenvalues

$$2.4069 \ldots \pm i \times 3.2110 \ldots$$

and the real eigenvalue $0.1863\ldots$. The standard (complex) Schur form is given by the pair of matrices

$$V = \begin{pmatrix} 0.3381 - 0.8462i & 0.3572 - 0.1071i & 0.1749 \\ 0.3193 - 0.0105i & -0.2263 - 0.6786i & -0.6214 \\ 0.1824 + 0.1852i & -0.2659 - 0.5277i & 0.7637 \end{pmatrix}$$

and

$$S = \begin{pmatrix} 2.4069 + 3.2110i & 4.6073 - 4.7030i & -2.3418 - 5.2330i \\ 0 & 2.4069 - 3.2110i & -2.0251 - 1.2016i \\ 0 & 0 & 0.1863 \end{pmatrix}.$$

It is possible to avoid complex arithmetic by using the quasi-Schur form which consists of the pair of matrices

$$U = \begin{pmatrix} -0.9768 & 0.1236 & 0.1749 \\ -0.0121 & 0.7834 & -0.6214 \\ 0.2138 & 0.6091 & 0.7637 \end{pmatrix}$$

and

$$R = \begin{pmatrix} 1.3129 & -7.7033 & 6.0407 \\ 1.4938 & 3.5008 & -1.3870 \\ 0 & 0 & 0.1863 \end{pmatrix}.$$

We conclude this section by pointing out that the Schur and the quasi-Schur forms of a given matrix are in no way unique. In addition to the dependence on the ordering of the eigenvalues, any column of Q can be multiplied by a complex sign $e^{i\theta}$ and a new corresponding R can be found. For the quasi-Schur form, there are infinitely many ways to select the 2×2 blocks, corresponding to applying arbitrary rotations to the columns of Q associated with these blocks.

1.8.4 APPLICATION TO POWERS OF MATRICES

The analysis of many numerical techniques is based on understanding the behavior of the successive powers A^k of a given matrix A. In this regard, the following theorem plays a fundamental role in numerical linear algebra, more particularly in the analysis of iterative methods.

THEOREM 1.4 *The sequence A^k, $k = 0, 1, \ldots$, converges to zero if and only if $\rho(A) < 1$.*

Proof. To prove the necessary condition, assume that $A^k \to 0$ and consider u_1 a unit eigenvector associated with an eigenvalue λ_1 of maximum modulus. We have

$$A^k u_1 = \lambda_1^k u_1,$$

which implies, by taking the 2-norms of both sides,

$$|\lambda_1^k| = \|A^k u_1\|_2 \to 0.$$

This shows that $\rho(A) = |\lambda_1| < 1$.

The Jordan canonical form must be used to show the sufficient condition. Assume that $\rho(A) < 1$. Start with the equality

$$A^k = XJ^kX^{-1}.$$

To prove that A^k converges to zero, it is sufficient to show that J^k converges to zero. An important observation is that J^k preserves its block form. Therefore, it is sufficient to prove that each of the Jordan blocks converges to zero. Each block is of the form

$$J_i = \lambda_i I + E_i$$

where E_i is a nilpotent matrix of index l_i, i.e., $E_i^{l_1} = 0$. Therefore, for $k \geq l_i$,

$$J_i^k = \sum_{j=0}^{l_1-1} \frac{k!}{j!(k-j)!} \lambda_i^{k-j} E_i^j.$$

Using the triangle inequality for any norm and taking $k \geq l_i$ yields

$$\|J_i^k\| \leq \sum_{j=0}^{l_1-1} \frac{k!}{j!(k-j)!} |\lambda_i|^{k-j} \|E_i^j\|.$$

Since $|\lambda_i| < 1$, each of the terms in this *finite* sum converges to zero as $k \to \infty$. Therefore, the matrix J_i^k converges to zero. ■

An equally important result is stated in the following theorem.

THEOREM 1.5 *The series*

$$\sum_{k=0}^{\infty} A^k$$

converges if and only if $\rho(A) < 1$. Under this condition, $I - A$ is nonsingular and the limit of the series is equal to $(I - A)^{-1}$.

Proof. The first part of the theorem is an immediate consequence of Theorem 1.4. Indeed, if the series converges, then $\|A^k\| \to 0$. By the previous theorem, this implies that $\rho(A) < 1$. To show that the converse is also true, use the equality

$$I - A^{k+1} = (I - A)(I + A + A^2 + \ldots + A^k)$$

and exploit the fact that since $\rho(A) < 1$, then $I - A$ is nonsingular, and therefore,

$$(I - A)^{-1}(I - A^{k+1}) = I + A + A^2 + \ldots + A^k.$$

This shows that the series converges since the left-hand side will converge to $(I - A)^{-1}$. In addition, it also shows the second part of the theorem. ■

Another important consequence of the Jordan canonical form is a result that relates the spectral radius of a matrix to its matrix norm.

THEOREM 1.6 *For any matrix norm $\|.\|$, we have*

$$\lim_{k\to\infty} \|A^k\|^{1/k} = \rho(A).$$

Proof. The proof is a direct application of the Jordan canonical form and is the subject of Exercise 10. ∎

1.9 NORMAL AND HERMITIAN MATRICES

This section examines specific properties of normal matrices and Hermitian matrices, including some optimality properties related to their spectra. The most common normal matrices that arise in practice are Hermitian or skew-Hermitian.

1.9.1 NORMAL MATRICES

By definition, a matrix is said to be normal if it commutes with its transpose conjugate, i.e., if it satisfies the relation

$$A^H A = AA^H. \tag{1.25}$$

An immediate property of normal matrices is stated in the following lemma.

LEMMA 1.1 *If a normal matrix is triangular, then it is a diagonal matrix.*

Proof. Assume, for example, that A is upper triangular and normal. Compare the first diagonal element of the left-hand side matrix of (1.25) with the corresponding element of the matrix on the right-hand side. We obtain that

$$|a_{11}|^2 = \sum_{j=1}^{n} |a_{1j}|^2,$$

which shows that the elements of the first row are zeros except for the diagonal one. The same argument can now be used for the second row, the third row, and so on to the last row, to show that $a_{ij} = 0$ for $i \neq j$. ∎

A consequence of this lemma is the following important result.

THEOREM 1.7 *A matrix is normal if and only if it is unitarily similar to a diagonal matrix.*

Proof. It is straightforward to verify that a matrix which is unitarily similar to a diagonal matrix is normal. We now prove that any normal matrix A is unitarily similar to a diagonal matrix. Let $A = QRQ^H$ be the Schur canonical form of A where Q is unitary and R is upper triangular. By the normality of A,

$$QR^HQ^HQRQ^H = QRQ^HQR^HQ^H$$

or,

$$QR^HRQ^H = QRR^HQ^H.$$

Upon multiplication by Q^H on the left and Q on the right, this leads to the equality $R^HR = RR^H$ which means that R is normal, and according to the previous lemma this is only possible if R is diagonal. ■

Thus, any normal matrix is diagonalizable and admits an orthonormal basis of eigenvectors, namely, the column vectors of Q.

The following result will be used in a later chapter. The question that is asked is: Assuming that any eigenvector of a matrix A is also an eigenvector of A^H, is A normal? If A had a full set of eigenvectors, then the result is true and easy to prove. Indeed, if V is the $n \times n$ matrix of common eigenvectors, then $AV = VD_1$ and $A^HV = VD_2$, with D_1 and D_2 diagonal. Then, $AA^HV = VD_1D_2$ and $A^HAV = VD_2D_1$ and, therefore, $AA^H = A^HA$. It turns out that the result is true in general, i.e., independently of the number of eigenvectors that A admits.

LEMMA 1.2 *A matrix A is normal if and only if each of its eigenvectors is also an eigenvector of A^H.*

Proof. If A is normal, then its left and right eigenvectors are identical, so the sufficient condition is trivial. Assume now that a matrix A is such that each of its eigenvectors $v_i, i = 1, \dots, k$, with $k \leq n$ is an eigenvector of A^H. For each eigenvector v_i of A, $Av_i = \lambda_i v_i$, and since v_i is also an eigenvector of A^H, then $A^H v_i = \mu v_i$. Observe that $(A^H v_i, v_i) = \mu(v_i, v_i)$ and because $(A^H v_i, v_i) = (v_i, Av_i) = \bar{\lambda}_i(v_i, v_i)$, it follows that $\mu = \bar{\lambda}_i$. Next, it is proved by contradiction that there are no elementary divisors. Assume that the contrary is true for λ_i. Then, the first principal vector u_i associated with λ_i is defined by

$$(A - \lambda_i I)u_i = v_i.$$

Taking the inner product of the above relation with v_i, we obtain

$$(Au_i, v_i) = \lambda_i(u_i, v_i) + (v_i, v_i). \tag{1.26}$$

On the other hand, it is also true that

$$(Au_i, v_i) = (u_i, A^H v_i) = (u_i, \bar{\lambda}_i v_i) = \lambda_i(u_i, v_i). \tag{1.27}$$

A result of (1.26) and (1.27) is that $(v_i, v_i) = 0$ which is a contradiction. Therefore, A has a full set of eigenvectors. This leads to the situation discussed just before the lemma, from which it is concluded that A must be normal. ■

Clearly, Hermitian matrices are a particular case of normal matrices. Since a normal matrix satisfies the relation $A = QDQ^H$, with D diagonal and Q unitary, the eigenvalues of A are the diagonal entries of D. Therefore, if these entries are real it is clear that $A^H = A$. This is restated in the following corollary.

COROLLARY 1.1 *A normal matrix whose eigenvalues are real is Hermitian.*

As will be seen shortly, the converse is also true, i.e., a Hermitian matrix has real eigenvalues.

An eigenvalue λ of any matrix satisfies the relation

$$\lambda = \frac{(Au, u)}{(u, u)},$$

where u is an associated eigenvector. Generally, one might consider the complex scalars

$$\mu(x) = \frac{(Ax, x)}{(x, x)}, \tag{1.28}$$

defined for any nonzero vector in $\mathbb{C}^n$. These ratios are known as *Rayleigh quotients* and are important both for theoretical and practical purposes. The set of all possible Rayleigh quotients as x runs over $\mathbb{C}^n$ is called the *field of values* of A. This set is clearly bounded since each $|\mu(x)|$ is bounded by the the 2-norm of A, i.e., $|\mu(x)| \leq \|A\|_2$ for all x.

If a matrix is normal, then any vector x in $\mathbb{C}^n$ can be expressed as

$$\sum_{i=1}^{n} \xi_i q_i,$$

where the vectors q_i form an orthogonal basis of eigenvectors, and the expression for $\mu(x)$ becomes

$$\mu(x) \;=\; \frac{(Ax, x)}{(x, x)} \;=\; \frac{\sum_{k=1}^{n} \lambda_k |\xi_k|^2}{\sum_{k=1}^{n} |\xi_k|^2} \;\equiv\; \sum_{k=1}^{n} \beta_k \lambda_k, \tag{1.29}$$

where

$$0 \leq \beta_i = \frac{|\xi_i|^2}{\sum_{k=1}^{n} |\xi_k|^2} \leq 1, \quad \text{and} \quad \sum_{i=1}^{n} \beta_i = 1.$$

From a well known characterization of convex hulls established by Hausdorff (Hausdorff's convex hull theorem), this means that the set of all possible Rayleigh quotients as x runs over all of $\mathbb{C}^n$ is equal to the convex hull of the λ_i's. This leads to the following theorem which is stated without proof.

THEOREM 1.8 *The field of values of a normal matrix is equal to the convex hull of its spectrum.*

The next question is whether or not this is also true for nonnormal matrices and the answer is no: The convex hull of the eigenvalues and the field of values of a nonnormal matrix are different in general. As a generic example, one can take any nonsymmetric real matrix which has real eigenvalues only. In this case, the convex hull of the spectrum is a real interval but its field of values will contain imaginary values. See Exercise 12 for another example. It has been shown (Hausdorff) that the field of values of a matrix is a convex set. Since the eigenvalues are members of the field of values, their convex hull is contained in the field of values. This is summarized in the following proposition.

PROPOSITION 1.5 *The field of values of an arbitrary matrix is a convex set which contains the convex hull of its spectrum. It is equal to the convex hull of the spectrum when the matrix is normal.*

1.9.2 HERMITIAN MATRICES

A first result on Hermitian matrices is the following.

THEOREM 1.9 *The eigenvalues of a Hermitian matrix are real, i.e.,* $\sigma(A) \subset \mathbb{R}$.

Proof. Let λ be an eigenvalue of A and u an associated eigenvector or 2-norm unity. Then

$$\lambda = (Au, u) = (u, Au) = \overline{(Au, u)} = \overline{\lambda},$$

which is the stated result. ∎

It is not difficult to see that if, in addition, the matrix is real, then the eigenvectors can be chosen to be real; see Exercise 21. Since a Hermitian matrix is normal, the following is a consequence of Theorem 1.7.

THEOREM 1.10 *Any Hermitian matrix is unitarily similar to a real diagonal matrix.*

In particular a Hermitian matrix admits a set of orthonormal eigenvectors that form a basis of $\mathbb{C}^n$.

In the proof of Theorem 1.8 we used the fact that the inner products (Au, u) are real. Generally, it is clear that any Hermitian matrix is such that (Ax, x) is real for any vector $x \in \mathbb{C}^n$. It turns out that the converse is also true, i.e., it can be shown that if (Az, z) is real for all vectors z in $\mathbb{C}^n$, then the matrix A is Hermitian; see Exercise 15.

Eigenvalues of Hermitian matrices can be characterized by optimality properties of the Rayleigh quotients (1.28). The best known of these is the min-max principle. We now label all the eigenvalues of A in descending order:

$$\lambda_1 \;\geq\; \lambda_2 \;\geq\; \cdots \;\geq\; \lambda_n.$$

Here, the eigenvalues are not necessarily distinct and they are repeated, each according to its multiplicity. In the following theorem, known as the *Min-Max Theorem*, S represents a generic subspace of $\mathbb{C}^n$.

THEOREM 1.11 *The eigenvalues of a Hermitian matrix* A *are characterized by the relation*

$$\lambda_k = \min_{S,\, \dim\, (S)=n-k+1} \quad \max_{x \in S, x \neq 0} \frac{(Ax, x)}{(x, x)}. \tag{1.30}$$

Proof. Let $\{q_i\}_{i=1,\ldots,n}$ be an orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of A associated with $\lambda_1, \ldots, \lambda_n$ respectively. Let S_k be the subspace spanned by the first k of these vectors and denote by $\mu(S)$ the maximum of $(Ax,x)/(x,x)$ over all nonzero vectors of a subspace S. Since the dimension of S_k is k, a well known theorem of linear algebra shows that its intersection with any subspace S of dimension $n-k+1$ is not reduced to $\{0\}$, i.e., there is vector x in $S \bigcap S_k$. For this $x = \sum_{i=1}^{k} \xi_i q_i$, we have

$$\frac{(Ax,x)}{(x,x)} \;=\; \frac{\sum_{i=1}^{k} \lambda_i |\xi_i|^2}{\sum_{i=1}^{k} |\xi_i|^2} \;\geq\; \lambda_k$$

so that $\mu(S) \;\geq\; \lambda_k$.

Consider, on the other hand, the particular subspace S_0 of dimension $n-k+1$ which is spanned by $q_k, \ldots, q_n$. For each vector x in this subspace, we have

$$\frac{(Ax,x)}{(x,x)} = \frac{\sum_{i=k}^{n} \lambda_i |\xi_i|^2}{\sum_{i=k}^{n} |\xi_i|^2} \;\leq\; \lambda_k$$

so that $\mu(S_0) \;\leq\; \lambda_k$. In other words, as S runs over all the $(n-k+1)$-dimensional subspaces, $\mu(S)$ is always $\geq \lambda_k$ and there is at least one subspace S_0 for which $\mu(S_0) \leq \lambda_k$. This shows the desired result. ∎

The above result is often called the Courant-Fisher min-max principle or theorem. As a particular case, the largest eigenvalue of A satisfies

$$\lambda_1 = \max_{x \neq 0} \frac{(Ax,x)}{(x,x)}. \tag{1.31}$$

Actually, there are four different ways of rewriting the above characterization. The second formulation is

$$\lambda_k = \max_{S,\ \dim\ (S)=k} \ \min_{x \in S, x \neq 0} \frac{(Ax,x)}{(x,x)} \tag{1.32}$$

and the two other ones can be obtained from (1.30) and (1.32) by simply relabeling the eigenvalues increasingly instead of decreasingly. Thus, with our labeling of the eigenvalues in descending order, (1.32) tells us that the smallest eigenvalue satisfies

$$\lambda_n = \min_{x \neq 0} \frac{(Ax,x)}{(x,x)}, \tag{1.33}$$

with λ_n replaced by λ_1 if the eigenvalues are relabeled increasingly.

In order for all the eigenvalues of a Hermitian matrix to be positive, it is necessary and sufficient that

$$(Ax,x) > 0, \quad \forall\, x \in \mathbb{C}^n, \quad x \neq 0.$$

Such a matrix is called *positive definite*. A matrix which satisfies $(Ax,x) \geq 0$ for any x is said to be *positive semidefinite*. In particular, the matrix $A^H A$ is semipositive definite for any rectangular matrix, since

$$(A^H Ax, x) = (Ax, Ax) \;\geq\; 0, \quad \forall\, x.$$

Similarly, AA^H is also a Hermitian semipositive definite matrix. The square roots of the eigenvalues of $A^H A$ for a general rectangular matrix A are called the *singular values* of

A and are denoted by σ_i. In Section 1.5, we have stated without proof that the 2-norm of any matrix A is equal to the largest singular value σ_1 of A. This is now an obvious fact, because

$$\|A\|_2^2 = \max_{x\neq 0} \frac{\|Ax\|_2^2}{\|x\|_2^2} = \max_{x\neq 0} \frac{(Ax, Ax)}{(x,x)} = \max_{x\neq 0} \frac{(A^H Ax, x)}{(x,x)} = \sigma_1^2$$

which results from (1.31).

Another characterization of eigenvalues, known as the Courant characterization, is stated in the next theorem. In contrast with the min-max theorem, this property is recursive in nature.

THEOREM 1.12 *The eigenvalue λ_i and the corresponding eigenvector q_i of a Hermitian matrix are such that*

$$\lambda_1 = \frac{(Aq_1, q_1)}{(q_1, q_1)} = \max_{x\in\mathbb{C}^n, x\neq 0} \frac{(Ax,x)}{(x,x)}$$

and for $k > 1$,

$$\lambda_k = \frac{(Aq_k, q_k)}{(q_k, q_k)} = \max_{x\neq 0, q_1^H x = \ldots = q_{k-1}^H x = 0} \frac{(Ax,x)}{(x,x)}. \tag{1.34}$$

In other words, the maximum of the Rayleigh quotient over a subspace that is orthogonal to the first $k-1$ eigenvectors is equal to λ_k and is achieved for the eigenvector q_k associated with λ_k. The proof follows easily from the expansion (1.29) of the Rayleigh quotient.

1.10 NONNEGATIVE MATRICES, M-MATRICES

Nonnegative matrices play a crucial role in the theory of matrices. They are important in the study of convergence of iterative methods and arise in many applications including economics, queuing theory, and chemical engineering.

A *nonnegative matrix* is simply a matrix whose entries are nonnegative. More generally, a partial order relation can be defined on the set of matrices.

DEFINITION 1.3 *Let A and B be two $n\times m$ matrices. Then*

$$A \le B$$

if by definition, $a_{ij} \le b_{ij}$ for $1 \le i \le n$, $1 \le j \le m$. If O denotes the $n\times m$ zero matrix, then A is nonnegative *if $A \ge O$, and* positive *if $A > O$. Similar definitions hold in which "positive" is replaced by "negative".*

The binary relation "$\le$" imposes only a *partial* order on $\mathbb{R}^{n\times m}$ since two arbitrary matrices in $\mathbb{R}^{n\times m}$ are not necessarily comparable by this relation. For the remainder of this section, we now assume that only square matrices are involved. The next proposition lists a number of rather trivial properties regarding the partial order relation just defined.

PROPOSITION 1.6 *The following properties hold.*

1. *The relation $\leq$ for matrices is reflexive ($A \leq A$), antisymmetric (if $A \leq B$ and $B \leq A$, then $A = B$), and transitive (if $A \leq B$ and $B \leq C$, then $A \leq C$).*
2. *If A and B are nonnegative, then so is their product AB and their sum $A + B$.*
3. *If A is nonnegative, then so is A^k.*
4. *If $A \leq B$, then $A^T \leq B^T$.*
5. *If $O \leq A \leq B$, then $\|A\|_1 \leq \|B\|_1$ and similarly $\|A\|_\infty \leq \|B\|_\infty$.*

The proof of these properties is left as Exercise 23.

A matrix is said to be *reducible* if there is a permutation matrix P such that PAP^T is block upper triangular. Otherwise, it is *irreducible*. An important result concerning nonnegative matrices is the following theorem known as the Perron-Frobenius theorem.

THEOREM 1.13 *Let A be a real $n \times n$ nonnegative irreducible matrix. Then $\lambda \equiv \rho(A)$, the spectral radius of A, is a simple eigenvalue of A. Moreover, there exists an eigenvector u with positive elements associated with this eigenvalue.*

A relaxed version of this theorem allows the matrix to be reducible but the conclusion is somewhat weakened in the sense that the elements of the eigenvectors are only guaranteed to be *nonnegative*.

Next, a useful property is established.

PROPOSITION 1.7 *Let A, B, C be nonnegative matrices, with $A \leq B$. Then*

$$AC \leq BC \quad \text{and} \quad CA \leq CB.$$

Proof. Consider the first inequality only, since the proof for the second is identical. The result that is claimed translates into

$$\sum_{k=1}^{n} a_{ik} c_{kj} \leq \sum_{k=1}^{n} b_{ik} c_{kj}, \quad 1 \leq i, j \leq n,$$

which is clearly true by the assumptions. ∎

A consequence of the proposition is the following corollary.

COROLLARY 1.2 *Let A and B be two nonnegative matrices, with $A \leq B$. Then*

$$A^k \leq B^k, \quad \forall\, k \geq 0. \tag{1.35}$$

Proof. The proof is by induction. The inequality is clearly true for $k = 0$. Assume that (1.35) is true for k. According to the previous proposition, multiplying (1.35) from the left by A results in

$$A^{k+1} \leq AB^k. \tag{1.36}$$

Now, it is clear that if $B \geq 0$, then also $B^k \geq 0$, by Proposition 1.6. We now multiply both sides of the inequality $A \leq B$ by B^k to the right, and obtain

$$AB^k \leq B^{k+1}. \tag{1.37}$$

The inequalities (1.36) and (1.37) show that $A^{k+1} \leq B^{k+1}$, which completes the induction proof. ∎

A theorem which has important consequences on the analysis of iterative methods will now be stated.

THEOREM 1.14 *Let A and B be two square matrices that satisfy the inequalities*

$$O \leq A \leq B. \tag{1.38}$$

Then

$$\rho(A) \leq \rho(B). \tag{1.39}$$

Proof. The proof is based on the following equality stated in Theorem 1.6

$$\rho(X) = \lim_{k \to \infty} \|X^k\|^{1/k}$$

for any matrix norm. Choosing the $1-$norm, for example, we have from the last property in Proposition 1.6

$$\rho(A) = \lim_{k \to \infty} \|A^k\|_1^{1/k} \leq \lim_{k \to \infty} \|B^k\|_1^{1/k} = \rho(B)$$

which completes the proof. ∎

THEOREM 1.15 *Let B be a nonnegative matrix. Then $\rho(B) < 1$ if and only if $I - B$ is nonsingular and $(I - B)^{-1}$ is nonnegative.*

Proof. Define $C = I - B$. If it is assumed that $\rho(B) < 1$, then by Theorem 1.5, $C = I - B$ is nonsingular and

$$C^{-1} = (I - B)^{-1} = \sum_{i=0}^{\infty} B^i. \tag{1.40}$$

In addition, since $B \geq 0$, all the powers of B as well as their sum in (1.40) are also nonnegative.

To prove the sufficient condition, assume that C is nonsingular and that its inverse is nonnegative. By the Perron-Frobenius theorem, there is a nonnegative eigenvector u associated with $\rho(B)$, which is an eigenvalue, i.e.,

$$Bu = \rho(B)u$$

or, equivalently,

$$C^{-1}u = \frac{1}{1 - \rho(B)}u.$$

Since u and C^{-1} are nonnegative, and $I - B$ is nonsingular, this shows that $1 - \rho(B) > 0$, which is the desired result. ∎

DEFINITION 1.4 *A matrix is said to be an M-matrix if it satisfies the following four properties:*

1. $a_{i,i} > 0$ *for* $i = 1, \ldots, n$.
2. $a_{i,j} \leq 0$ *for* $i \neq j,\ i, j = 1, \ldots, n$.
3. A *is nonsingular.*
4. $A^{-1} \geq 0$.

In reality, the four conditions in the above definition are somewhat redundant and equivalent conditions that are more rigorous will be given later. Let A be any matrix which satisfies properties (1) and (2) in the above definition and let D be the diagonal of A. Since $D > 0$,

$$A = D - (D - A) = D\left(I - (I - D^{-1}A)\right).$$

Now define

$$B \equiv I - D^{-1}A.$$

Using the previous theorem, $I - B = D^{-1}A$ is nonsingular and $(I - B)^{-1} = A^{-1}D \geq 0$ if and only if $\rho(B) < 1$. It is now easy to see that conditions (3) and (4) of Definition 1.4 can be replaced by the condition $\rho(B) < 1$.

THEOREM 1.16 *Let a matrix A be given such that*

1. $a_{i,i} > 0$ *for* $i = 1, \ldots, n$.
2. $a_{i,j} \leq 0$ *for* $i \neq j,\ i, j = 1, \ldots, n$.

Then A is an M-matrix if and only if

3. $\rho(B) < 1$, *where* $B = I - D^{-1}A$.

Proof. From the above argument, an immediate application of Theorem 1.15 shows that properties (3) and (4) of the above definition are equivalent to $\rho(B) < 1$, where $B = I - C$ and $C = D^{-1}A$. In addition, C is nonsingular iff A is and C^{-1} is nonnegative iff A is. ∎

The next theorem shows that the condition (1) in Definition 1.4 is implied by the other three.

THEOREM 1.17 *Let a matrix A be given such that*

1. $a_{i,j} \leq 0$ *for* $i \neq j,\ i, j = 1, \ldots, n$.
2. A *is nonsingular.*
3. $A^{-1} \geq 0$.

Then

4. $a_{i,i} > 0$ *for* $i = 1, \ldots, n$, *i.e.,* A *is an* M-*matrix.*
5. $\rho(B) < 1$ *where* $B = I - D^{-1}A$.

Proof. Define $C \equiv A^{-1}$. Writing that $(AC)_{ii} = 1$ yields

$$\sum_{k=1}^{n} a_{ik} c_{ki} = 1$$

which gives

$$a_{ii} c_{ii} = 1 - \sum_{\substack{k=1 \\ k \neq i}}^{n} a_{ik} c_{ki}.$$

Since $a_{ik} c_{ki} \leq 0$ for all k, the right-hand side is ≥ 1 and since $c_{ii} \geq 0$, then $a_{ii} > 0$. The second part of the result now follows immediately from an application of the previous theorem. ■

Finally, this useful result follows.

THEOREM 1.18 *Let A, B be two matrices which satisfy*

1. $A \leq B$.
2. $b_{ij} \leq 0$ *for all* $i \neq j$.

Then if A is an M-matrix, so is the matrix B.

Proof. Assume that A is an M-matrix and let D_X denote the diagonal of a matrix X. The matrix D_B is positive because

$$D_B \geq D_A > 0.$$

Consider now the matrix $I - D_B^{-1} B$. Since $A \leq B$, then

$$D_A - A \geq D_B - B \geq O$$

which, upon multiplying through by D_A^{-1}, yields

$$I - D_A^{-1} A \geq D_A^{-1} (D_B - B) \geq D_B^{-1} (D_B - B) = I - D_B^{-1} B \geq O.$$

Since the matrices $I - D_B^{-1} B$ and $I - D_A^{-1} A$ are nonnegative, Theorems 1.14 and 1.16 imply that

$$\rho(I - D_B^{-1} B) \leq \rho(I - D_A^{-1} A) < 1.$$

This establishes the result by using Theorem 1.16 once again. ■

1.11 POSITIVE-DEFINITE MATRICES

A *real* matrix is said to be *positive definite* or *positive real* if

$$(Au, u) > 0, \quad \forall\, u \in \mathbb{R}^n,\ u \neq 0. \tag{1.41}$$

It must be emphasized that this definition is only useful when formulated entirely for real variables. Indeed, if u were not restricted to be real, then assuming that (Au, u) is real for all u complex would imply that A is Hermitian; see Exercise 15. If, in addition to Definition 1.41, A is symmetric (real), then A is said to be *Symmetric Positive Definite* (SPD). Similarly, if A is Hermitian, then A is said to be *Hermitian Positive Definite* (HPD). Some properties of HPD matrices were seen in Section 1.9, in particular with regards to their eigenvalues. Now the more general case where A is non-Hermitian and positive definite is considered.

We begin with the observation that any square matrix (real or complex) can be decomposed as

$$A = H + iS, \tag{1.42}$$

in which

$$H = \frac{1}{2}(A + A^H) \tag{1.43}$$

$$S = \frac{1}{2i}(A - A^H). \tag{1.44}$$

Note that both H and S are Hermitian while the matrix iS in the decomposition (1.42) is skew-Hermitian. The matrix H in the decomposition is called the *Hermitian part* of A, while the matrix iS is the *skew-Hermitian part* of A. The above decomposition is the analogue of the decomposition of a complex number z into $z = x + iy$,

$$x = \Re e(z) = \frac{1}{2}(z + \bar{z}),\ y = \Im m(z) = \frac{1}{2i}(z - \bar{z}).$$

When A is real and u is a real vector then (Au, u) is real and, as a result, the decomposition (1.42) immediately gives the equality

$$(Au, u) = (Hu, u). \tag{1.45}$$

This results in the following theorem.

THEOREM 1.19 *Let A be a real positive definite matrix. Then A is nonsingular. In addition, there exists a scalar $\alpha > 0$ such that*

$$(Au, u) \geq \alpha \|u\|_2^2, \tag{1.46}$$

for any real vector u.

Proof. The first statement is an immediate consequence of the definition of positive definiteness. Indeed, if A were singular, then there would be a nonzero vector such that $Au = 0$ and as a result $(Au, u) = 0$ for this vector, which would contradict (1.41). We now prove the second part of the theorem. From (1.45) and the fact that A is positive definite, we conclude that H is HPD. Hence, from (1.33) based on the min-max theorem, we get

$$\min_{u \neq 0} \frac{(Au, u)}{(u, u)} = \min_{u \neq 0} \frac{(Hu, u)}{(u, u)} \geq \lambda_{min}(H) > 0.$$

Taking $\alpha \equiv \lambda_{min}(H)$ yields the desired inequality (1.46). ∎

A simple yet important result which locates the eigenvalues of A in terms of the spectra of H and S can now be proved.

THEOREM 1.20 *Let A be any square (possibly complex) matrix and let $H = \frac{1}{2}(A + A^H)$ and $S = \frac{1}{2i}(A - A^H)$. Then any eigenvalue λ_j of A is such that*

$$\lambda_{min}(H) \le \Re e(\lambda_j) \le \lambda_{max}(H) \tag{1.47}$$

$$\lambda_{min}(S) \le \Im m(\lambda_j) \le \lambda_{max}(S). \tag{1.48}$$

Proof. When the decomposition (1.42) is applied to the Rayleigh quotient of the eigenvector u_j associated with λ_j, we obtain

$$\lambda_j = (Au_j, u_j) = (Hu_j, u_j) + i(Su_j, u_j), \tag{1.49}$$

assuming that $\|u_j\|_2 = 1$. This leads to

$$\Re e(\lambda_j) = (Hu_j, u_j)$$
$$\Im m(\lambda_j) = (Su_j, u_j).$$

The result follows using properties established in Section 1.9. ■

Thus, the eigenvalues of a matrix are contained in a rectangle defined by the eigenvalues of its Hermitian part and its non-Hermitian part. In the particular case where A is real, then iS is skew-Hermitian and its eigenvalues form a set that is symmetric with respect to the real axis in the complex plane. Indeed, in this case, iS is real and its eigenvalues come in conjugate pairs.

Note that all the arguments herein are based on the field of values and, therefore, they provide ways to localize the eigenvalues of A from knowledge of the field of values. However, this approximation can be inaccurate in some cases.

Example 1.3 Consider the matrix

$$A = \begin{pmatrix} 1 & 1 \\ 10^4 & 1 \end{pmatrix}.$$

The eigenvalues of A are -99 and 101. Those of H are $1 \pm (10^4 + 1)/2$ and those of iS are $\pm i(10^4 - 1)/2$.

When a matrix B is Symmetric Positive Definite, the mapping

$$x, y \quad \rightarrow \quad (x, y)_B \equiv (Bx, y) \tag{1.50}$$

from $\mathbb{C}^n \times \mathbb{C}^n$ to $\mathbb{C}$ is a proper inner product on $\mathbb{C}^n$, in the sense defined in Section 1.4. The associated norm is often referred to as the *energy norm*. Sometimes, it is possible to find an appropriate HPD matrix B which makes a given matrix A Hermitian, i.e., such that

$$(Ax, y)_B = (x, Ay)_B, \quad \forall\, x, y$$

although A is a non-Hermitian matrix with respect to the Euclidean inner product. The simplest examples are $A = B^{-1}C$ and $A = CB$, where C is Hermitian and B is Hermitian Positive Definite.

PROJECTION OPERATORS

1.12

Projection operators or *projectors* play an important role in numerical linear algebra, particularly in iterative methods for solving various matrix problems. This section introduces these operators from a purely algebraic point of view and gives a few of their important properties.

1.12.1 RANGE AND NULL SPACE OF A PROJECTOR

A projector P is any linear mapping from $\mathbb{C}^n$ to itself which is idempotent, i.e., such that

$$P^2 = P.$$

A few simple properties follow from this definition. First, if P is a projector, then so is $(I - P)$, and the following relation holds,

$$\mathrm{Ker}(P) = \mathrm{Ran}(I - P).$$

In addition, the two subspaces $\mathrm{Ker}(P)$ and $\mathrm{Ran}(P)$ intersect only at the element zero. Indeed, if a vector x belongs to $\mathrm{Ran}(P)$, then $Px = x$, by the idempotence property. If it is also in $\mathrm{Ker}(P)$, then $Px = 0$. Hence, $x = Px = 0$ which proves the result. Moreover, every element of $\mathbb{C}^n$ can be written as $x = Px + (I - P)x$. Therefore, the space $\mathbb{C}^n$ can be decomposed as the direct sum

$$\mathbb{C}^n = \mathrm{Ker}(P) \ \oplus \ \mathrm{Ran}(P).$$

Conversely, every pair of subspaces M and S which forms a direct sum of $\mathbb{C}^n$ defines a unique projector such that $\mathrm{Ran}(P) = M$ and $\mathrm{Ker}(P) = S$. This associated projector P maps an element x of $\mathbb{C}^n$ into the component x_1, where x_1 is the M-component in the unique decomposition $x = x_1 + x_2$ associated with the direct sum.

In fact, this association is unique, that is, an arbitrary projector P can be entirely determined by the given of two subspaces: (1) The range M of P, and (2) its null space S which is also the range of $I - P$. For any x, the vector Px satisfies the conditions,

$$\begin{aligned} Px &\in M \\ x - Px &\in S. \end{aligned}$$

The linear mapping P is said to project x *onto* M and *along* or *parallel to* the subspace S. If P is of rank m, then the range of $I - P$ is of dimension $n - m$. Therefore, it is natural to define S through its orthogonal complement $L = S^{\perp}$ which has dimension m. The above conditions that define $u = Px$ for any x become

$$u \in M \tag{1.51}$$

$$x - u \perp L. \tag{1.52}$$

These equations define a projector P *onto* M and *orthogonal* to the subspace L. The first statement, (1.51), establishes the m degrees of freedom, while the second, (1.52), gives the m constraints that define Px from these degrees of freedom. The general definition of projectors is illustrated in Figure 1.1.

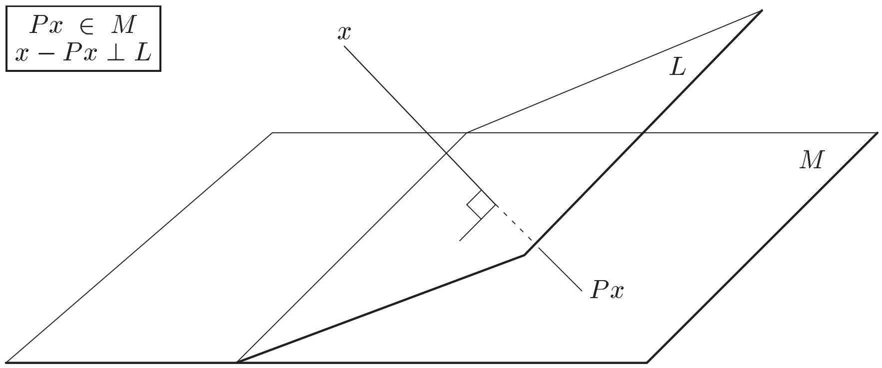

Figure 1.1 *Projection of x onto M and orthogonal to L.*

The question now is: Given two arbitrary subspaces, M and L both of dimension m, is it always possible to define a projector onto M orthogonal to L through the conditions (1.51) and (1.52)? The following lemma answers this question.

LEMMA 1.3 *Given two subspaces M and L of the same dimension m, the following two conditions are mathematically equivalent.*

- ***i.*** *No nonzero vector of M is orthogonal to L;*
- ***ii.*** *For any x in $\mathbb{C}^n$ there is a unique vector u which satisfies the conditions (1.51) and (1.52).*

Proof. The first condition states that any vector which is in M and also orthogonal to L must be the zero vector. It is equivalent to the condition

$$M \cap L^{\perp} = \{0\}.$$

Since L is of dimension m, $L^{\perp}$ is of dimension $n-m$ and the above condition is equivalent to the condition that

$$\mathbb{C}^n = M \oplus L^{\perp}. \tag{1.53}$$

This in turn is equivalent to the statement that for any x, there exists a unique pair of vectors u, w such that

$$x = u + w,$$

where u belongs to M, and $w = x - u$ belongs to $L^{\perp}$, a statement which is identical with ii. ∎

In summary, given two subspaces M and L, satisfying the condition $M \cap L^{\perp} = \{0\}$, there is a projector P onto M orthogonal to L, which defines the projected vector u of any vector

x from equations (1.51) and (1.52). This projector is such that

$$\mathrm{Ran}(P) = M, \qquad \mathrm{Ker}(P) = L^\perp.$$

In particular, the condition $Px = 0$ translates into $x \in \mathrm{Ker}(P)$ which means that $x \in L^\perp$. The converse is also true. Hence, the following useful property,

$$Px = 0 \quad \text{iff} \quad x \perp L. \tag{1.54}$$

1.12.2 MATRIX REPRESENTATIONS

Two bases are required to obtain a matrix representation of a general projector: a basis $V = [v_1, \ldots, v_m]$ for the subspace $M = \mathrm{Ran}(P)$ and a second one $W = [w_1, \ldots, w_m]$ for the subspace L. These two bases are *biorthogonal* when

$$(v_i, w_j) = \delta_{ij}. \tag{1.55}$$

In matrix form this means $W^H V = I$. Since Px belongs to M, let Vy be its representation in the V basis. The constraint $x - Px \perp L$ is equivalent to the condition,

$$((x - Vy), w_j) = 0 \quad \text{for } j = 1, \ldots, m.$$

In matrix form, this can be rewritten as

$$W^H (x - Vy) = 0. \tag{1.56}$$

If the two bases are biorthogonal, then it follows that $y = W^H x$. Therefore, in this case, $Px = VW^H x$, which yields the matrix representation of P,

$$P = VW^H. \tag{1.57}$$

In case the bases V and W are not biorthogonal, then it is easily seen from the condition (1.56) that

$$P = V(W^H V)^{-1} W^H. \tag{1.58}$$

If we assume that no vector of M is orthogonal to L, then it can be shown that the $m \times m$ matrix $W^H V$ is nonsingular.

1.12.3 ORTHOGONAL AND OBLIQUE PROJECTORS

An important class of projectors is obtained in the case when the subspace L is equal to M, i.e., when

$$\mathrm{Ker}(P) = \mathrm{Ran}(P)^\perp.$$

Then, the projector P is said to be the *orthogonal projector* onto M. A projector that is not orthogonal is *oblique*. Thus, an orthogonal projector is defined through the following requirements satisfied for any vector x,

$$Px \in M \quad \text{and} \quad (I - P)\, x \perp M \tag{1.59}$$

or equivalently,

$$Px \in M \quad \text{and} \quad ((I-P)x, y) = 0 \quad \forall\, y \in M.$$

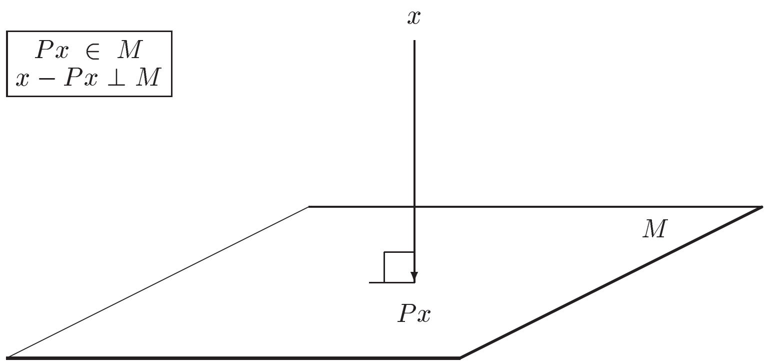

Figure 1.2 *Orthogonal projection of x onto a subspace M.*

It is interesting to consider the mapping P^H defined as the adjoint of P

$$(P^H x, y) = (x, Py), \quad \forall\, x,\ \forall\, y. \tag{1.60}$$

First note that P^H is also a projector because for all x and y,

$$((P^H)^2 x, y) = (P^H x, Py) = (x, P^2 y) = (x, Py) = (P^H x, y).$$

A consequence of the relation (1.60) is

$$\mathrm{Ker}(P^H) = \mathrm{Ran}(P)^\perp \tag{1.61}$$
$$\mathrm{Ker}(P) = \mathrm{Ran}(P^H)^\perp. \tag{1.62}$$

The above relations lead to the following proposition.

PROPOSITION 1.8 *A projector is orthogonal if and only if it is Hermitian.*

Proof. By definition, an orthogonal projector is one for which $\mathrm{Ker}(P) = \mathrm{Ran}(P)^\perp$. Therefore, by (1.61), if P is Hermitian, then it is orthogonal. Conversely, if P is orthogonal, then (1.61) implies $\mathrm{Ker}(P) = \mathrm{Ker}(P^H)$ while (1.62) implies $\mathrm{Ran}(P) = \mathrm{Ran}(P^H)$. Since P^H is a projector and since projectors are uniquely determined by their range and null spaces, this implies that $P = P^H$. ∎

Given *any* unitary $n \times m$ matrix V whose columns form an orthonormal basis of $M = \mathrm{Ran}(P)$, we can represent P by the matrix $P = VV^H$. This is a particular case of the matrix representation of projectors (1.57). In addition to being idempotent, the linear mapping associated with this matrix satisfies the characterization given above, i.e.,

$$VV^H x \ \in M \quad \text{and} \quad (I - VV^H)x \ \in \ M^\perp.$$

It is important to note that this representation of the orthogonal projector P is not unique. In fact, any orthonormal basis V will give a different representation of P in the above form. As a consequence for any two orthogonal bases V_1, V_2 of M, we must have $V_1 V_1^H = V_2 V_2^H$, an equality which can also be verified independently; see Exercise 26.

1.12.4 PROPERTIES OF ORTHOGONAL PROJECTORS

When P is an orthogonal projector, then the two vectors Px and $(I - P)x$ in the decomposition $x = Px + (I - P)x$ are orthogonal. The following relation results:

$$\|x\|_2^2 = \|Px\|_2^2 + \|(I - P)x\|_2^2.$$

A consequence of this is that for any x,

$$\|Px\|_2 \le \|x\|_2.$$

Thus, the maximum of $\|Px\|_2/\|x\|_2$, for all x in $\mathbb{C}^n$ does not exceed one. In addition the value one is reached for any element in $\mathrm{Ran}(P)$. Therefore,

$$\|P\|_2 = 1$$

for any orthogonal projector P.

An orthogonal projector has only two eigenvalues: zero or one. Any vector of the range of P is an eigenvector associated with the eigenvalue one. Any vector of the null-space is obviously an eigenvector associated with the eigenvalue zero.

Next, an important optimality property of orthogonal projectors is established.

THEOREM 1.21 *Let P be the orthogonal projector onto a subspace M. Then for any given vector x in $\mathbb{C}^n$, the following is true:*

$$\min_{y \in M} \|x - y\|_2 = \|x - Px\|_2. \tag{1.63}$$

Proof. Let y be any vector of M and consider the square of its distance from x. Since $x - Px$ is orthogonal to M to which $Px - y$ belongs, then

$$\|x - y\|_2^2 = \|x - Px + (Px - y)\|_2^2 = \|x - Px\|_2^2 + \|(Px - y)\|_2^2.$$

Therefore, $\|x - y\|_2 \ge \|x - Px\|_2$ for all y in M. This establishes the result by noticing that the minimum is reached for $y = Px$. ■

By expressing the conditions that define $y^* \equiv Px$ for an orthogonal projector P onto a subspace M, it is possible to reformulate the above result in the form of necessary and sufficient conditions which enable us to determine the best approximation to a given vector x in the least-squares sense.

COROLLARY 1.3 *Let a subspace M, and a vector x in $\mathbb{C}^n$ be given. Then*

$$\min_{y \in M} \|x - y\|_2 = \|x - y^*\|_2, \tag{1.64}$$

if and only if the following two conditions are satisfied,

$$\left\{ \begin{array}{lcl} y^* & \in & M \\ x - y^* & \perp & M. \end{array} \right.$$

1.13 BASIC CONCEPTS IN LINEAR SYSTEMS

Linear systems are among the most important and common problems encountered in scientific computing. From the theoretical point of view, the problem is rather easy and explicit solutions using determinants exist. In addition, it is well understood when a solution exists, when it does not, and when there are infinitely many solutions. However, the numerical viewpoint is far more complex. Approximations may be available but it may be difficult to estimate how accurate they are. This clearly will depend on the data at hand, i.e., primarily on the coefficient matrix. This section gives a very brief overview of the existence theory as well as the sensitivity of the solutions.

1.13.1 EXISTENCE OF A SOLUTION

Consider the *linear system*

$$Ax = b. \tag{1.65}$$

Here, x is termed the *unknown* and b the *right-hand side.* When solving the linear system (1.65), we distinguish three situations.

Case 1 The matrix A is nonsingular. There is a unique solution given by $x = A^{-1}b$.

Case 2 The matrix A is singular and $b \in \mathrm{Ran}(A)$. Since $b \in \mathrm{Ran}(A)$, there is an x_0 such that $Ax_0 = b$. Then $x_0 + v$ is also a solution for any v in $\mathrm{Ker}(A)$. Since $\mathrm{Ker}(A)$ is at least one-dimensional, there are infinitely many solutions.

Case 3 The matrix A is singular and $b \notin \mathrm{Ran}(A)$. There are no solutions.

Example 1.4 The simplest illustration of the above three cases is with small diagonal matrices. Let

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix} \quad b = \begin{pmatrix} 1 \\ 8 \end{pmatrix}.$$

Then A is nonsingular and there is a unique x given by

$$x = \begin{pmatrix} 0.5 \\ 2 \end{pmatrix}.$$

Now let

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Then A is singular and, as is easily seen, $b \in \mathrm{Ran}(A)$. For example, a particular element x_0 such that $Ax_0 = b$ is $x_0 = \binom{0.5}{0}$. The null space of A consists of all vectors whose first component is zero, i.e., all vectors of the form $\binom{0}{\alpha}$. Therefore, there are infinitely many solution which are given by

$$x(\alpha) = \begin{pmatrix} 0.5 \\ \alpha \end{pmatrix} \quad \forall\, \alpha.$$

Finally, let A be the same as in the previous case, but define the right-hand side as

$$b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

In this case there are no solutions because the second equation cannot be satisfied.

1.13.2 PERTURBATION ANALYSIS

Consider the linear system (1.65) where A is an $n \times n$ nonsingular matrix. Given any matrix E, the matrix $A(\epsilon) = A + \epsilon E$ is nonsingular for ϵ small enough, i.e., for $\epsilon \leq \alpha$ where α is some small number; see Exercise 32. Assume that we perturb the data in the above system, i.e., that we perturb the matrix A by ϵE and the right-hand side b by ϵe. The solution $x(\epsilon)$ of the perturbed system satisfies the equation,

$$(A + \epsilon E)x(\epsilon) = b + \epsilon e. \tag{1.66}$$

Let $\delta(\epsilon) = x(\epsilon) - x$. Then,

$$\begin{aligned} (A + \epsilon E)\delta(\epsilon) &= (b + \epsilon e) - (A + \epsilon E)x \\ &= \epsilon\ (e - Ex) \\ \delta(\epsilon) &= \epsilon\ (A + \epsilon E)^{-1}(e - Ex). \end{aligned}$$

As an immediate result, the function $x(\epsilon)$ is differentiable at $\epsilon = 0$ and its derivative is given by

$$x'(0) = \lim_{\epsilon \to 0} \frac{\delta(\epsilon)}{\epsilon} = A^{-1}\,(e - Ex)\,. \tag{1.67}$$

The size of the derivative of $x(\epsilon)$ is an indication of the size of the variation that the solution $x(\epsilon)$ undergoes when the data, i.e., the pair $[A, b]$ is perturbed in the direction $[E, e]$. In absolute terms, a small variation $[\epsilon E, \epsilon e]$ will cause the solution to vary by roughly $\epsilon x'(0) = \epsilon A^{-1}(e - Ex)$. The relative variation is such that

$$\frac{\|x(\epsilon) - x\|}{\|x\|} \leq \epsilon \|A^{-1}\| \left(\frac{\|e\|}{\|x\|} + \|E\| \right) + o(\epsilon).$$

Using the fact that $\|b\| \leq \|A\|\|x\|$ in the above equation yields

$$\frac{\|x(\epsilon) - x\|}{\|x\|} \leq \epsilon \|A\|\|A^{-1}\| \left(\frac{\|e\|}{\|b\|} + \frac{\|E\|}{\|A\|} \right) + o(\epsilon) \tag{1.68}$$

which relates the relative variation in the solution to the relative sizes of the perturbations. The quantity

$$\kappa(A) = \|A\|\,\|A^{-1}\|$$

is called the *condition number* of the linear system (1.65) with respect to the norm $\|.\|$. The condition number is relative to a norm. When using the standard norms $\|.\|_p, p = 1, \ldots, \infty$, it is customary to label $\kappa(A)$ with the same label as the associated norm. Thus,

$$\kappa_p(A) = \|A\|_p \|A^{-1}\|_p.$$

For large matrices, the determinant of a matrix is almost never a good indication of "near" singularity or degree of sensitivity of the linear system. The reason is that $\det(A)$ is the product of the eigenvalues which depends very much on a scaling of a matrix, whereas the condition number of a matrix is scaling-invariant. For example, for $A = \alpha I$ the determinant is $\det(A) = \alpha^n$, which can be very small if $|\alpha| < 1$, whereas $\kappa(A) = 1$ for any of the standard norms.

In addition, small eigenvalues do not always give a good indication of poor conditioning. Indeed, a matrix can have all its eigenvalues equal to one yet be poorly conditioned.

Example 1.5 The simplest example is provided by matrices of the form

$$A_n = I + \alpha e_1 e_n^T$$

for large α. The inverse of A_n is

$$A_n^{-1} = I - \alpha e_1 e_n^T$$

and for the ∞-norm we have

$$\|A_n\|_\infty = \|A_n^{-1}\|_\infty = 1 + |\alpha|$$

so that

$$\kappa_\infty(A_n) = (1 + |\alpha|)^2.$$

For a large α, this can give a very large condition number, whereas all the eigenvalues of A_n are equal to unity.

When an iterative procedure is used for solving a linear system, we typically face the problem of choosing a good stopping procedure for the algorithm. Often a residual norm,

$$\|r\| = \|b - A\tilde{x}\|$$

is available for some current approximation $\tilde{x}$ and an estimate of the absolute error $\|x - \tilde{x}\|$ or the relative error $\|x - \tilde{x}\|/\|x\|$ is desired. The following simple relation is helpful in this regard,

$$\frac{\|x - \tilde{x}\|}{\|x\|} \le \kappa(A) \, \frac{\|r\|}{\|b\|}.$$

It is necessary to have an estimate of the condition number $\kappa(A)$ in order to exploit the above relation.

EXERCISES

1. Verify that the Euclidean inner product defined by (1.4) does indeed satisfy the general definition of inner products on vector spaces.

2. Show that two eigenvectors associated with two distinct eigenvalues are linearly independent. In a more general sense, show that a family of eigenvectors associated with distinct eigenvalues forms a linearly independent family.

3. Show that if λ is any nonzero eigenvalue of the matrix AB, then it is also an eigenvalue of the matrix BA. Start with the particular case where A and B are square and B is nonsingular, then consider the more general case where A, B may be singular or even rectangular (but such that AB and BA are square).

4. Let A be an $n \times n$ orthogonal matrix, i.e., such that $A^H A = D$, where D is a diagonal matrix. Assuming that D is nonsingular, what is the inverse of A? Assuming that $D > 0$, how can A be transformed into a unitary matrix (by operations on its rows or columns)?

5. Show that the Frobenius norm is consistent. Can this norm be associated to two vector norms via (1.7)? What is the Frobenius norm of a diagonal matrix? What is the p-norm of a diagonal matrix (for any p)?

6. Find the Jordan canonical form of the matrix:

$$A = \begin{pmatrix} 1 & 2 & -4 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{pmatrix}.$$

Same question for the matrix obtained by replacing the element a_{33} by 1.

7. Give an alternative proof of Theorem 1.3 on the Schur form by starting from the Jordan canonical form. [Hint: Write $A = XJX^{-1}$ and use the QR decomposition of X.]

8. Show from the definition of determinants used in Section 1.2 that the characteristic polynomial is a polynomial of degree n for an $n \times n$ matrix.

9. Show that the characteristic polynomials of two similar matrices are equal.

10. Show that

$$\lim_{k \to \infty} \|A^k\|^{1/k} = \rho(A),$$

for any matrix norm. [Hint: Use the Jordan canonical form.]

11. Let X be a nonsingular matrix and, for any matrix norm $\|.\|$, define $\|A\|_X = \|AX\|$. Show that this is indeed a matrix norm. Is this matrix norm consistent? Show the same for $\|XA\|$ and $\|YAX\|$ where Y is also a nonsingular matrix. These norms are not, in general, associated with any vector norms, i.e., they can't be defined by a formula of the form (1.7). Why? What about the particular case $\|A\|' = \|XAX^{-1}\|$?

12. Find the field of values of the matrix

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

and verify that it is not equal to the convex hull of its eigenvalues.

13. Show that for a skew-Hermitian matrix S,

$$\Re e(Sx, x) = 0 \quad \text{for any } x \in \mathbb{C}^n.$$

14. Given an arbitrary matrix S, show that if $(Sx, x) = 0$ for all x in $\mathbb{C}^n$, then it is true that

$$(Sy, z) + (Sz, y) = 0 \quad \forall\, y,\ z\ \in\ \mathbb{C}^n.$$

[Hint: Expand $(S(y + z), y + z)$.]

15. Using the results of the previous two problems, show that if (Ax, x) is real for all x in $\mathbb{C}^n$, then A must be Hermitian. Would this result be true if the assumption were to be replaced by: (Ax, x) *is real for all real* x? Explain.

16. The definition of a positive definite matrix is that (Ax, x) be real and positive for all real vectors x. Show that this is equivalent to requiring that the Hermitian part of A, namely, $\frac{1}{2}(A + A^H)$, be (Hermitian) positive definite.

17. Let $A_1 = B^{-1}C$ and $A_2 = CB$ where C is a Hermitian matrix and B is a Hermitian Positive Definite matrix. Are A_1 and A_2 Hermitian *in general*? Show that A_1 and A_2 are Hermitian (self-adjoint) with respect to the B-inner product.

18. Let a matrix A be such that $A^H = p(A)$ where p is a polynomial. Show that A is normal. [Hint: Use Lemma 1.2.]

19. Show that A is normal iff its Hermitian and skew-Hermitian parts, as defined in Section 1.11, commute.

20. Let A be a Hermitian matrix and B a Hermitian Positive Definite matrix defining a B-inner product. Show that A is Hermitian (self-adjoint) with respect to the B-inner product if and only if A and B commute. What condition must satisfy B for the same condition to hold in the more general case where A is not Hermitian?

21. Let A be a real symmetric matrix and λ an eigenvalue of A. Show that if u is an eigenvector associated with λ, then so is $\bar{u}$. As a result, prove that for any eigenvalue of a real symmetric matrix, there is an associated eigenvector which is real.

22. Show that a Hessenberg matrix H such that $h_{j+1,j} \neq 0, j = 1, 2, \ldots, n - 1$, cannot be derogatory.

23. Prove all the properties listed in Proposition 1.6.

24. Let A be an M-matrix and u, v two nonnegative vectors such that $v^T A^{-1} u < 1$. Show that $A - uv^T$ is an M-matrix.

25. Show that if $O \leq A \leq B$ then $O \leq A^T A \leq B^T B$. Conclude that under the same assumption, we have $\|A\|_2 \leq \|B\|_2$.

26. Show that for two orthogonal bases V_1, V_2 of the same subspace M of $\mathbb{C}^n$ we have $V_1 V_1^H x = V_2 V_2^H x,\ \forall\, x$.

27. What are the eigenvalues of a projector? What about its eigenvectors?

28. Show that if two projectors P_1 and P_2 commute, then their product $P = P_1 P_2$ is a projector. What are the range and kernel of P?

29. Consider the matrix A of size $n \times n$ and the vector $x \in \mathbb{R}^n$,

$$A = \begin{pmatrix} 1 & -1 & -1 & -1 & \dots & -1 \\ 0 & 1 & -1 & -1 & \dots & -1 \\ 0 & 0 & 1 & -1 & \dots & -1 \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix} \qquad x = \begin{pmatrix} 1 \\ 1/2 \\ 1/4 \\ 1/8 \\ \vdots \\ 1/2^{n-1} \end{pmatrix}.$$

a. Compute Ax, $\|Ax\|_2$, and $\|x\|_2$.

b. Show that $\|A\|_2 \geq \sqrt{n}$.

c. Give a lower bound for $\kappa_2(A)$.

30. What is the inverse of the matrix A of the previous exercise? Give an expression of $\kappa_1(A)$ and $\kappa_\infty(A)$ based on this.

31. Find a small rank-one perturbation which makes the matrix A in Exercise 29 singular. Derive a lower bound for the singular values of A.

32. Consider a nonsingular matrix A. Given any matrix E, show that there exists α such that the matrix $A(\epsilon) = A + \epsilon E$ is nonsingular for all $\epsilon < \alpha$. What is the largest possible value for α satisfying the condition? [Hint: Consider the eigenvalues of the generalized eigenvalue problem $Au = \lambda Eu$.]

NOTES AND REFERENCES. For additional reading on the material presented in this chapter, see Golub and Van Loan [108], Datta [64], Stewart [202], and Varga [213]. Details on matrix eigenvalue problems can be found in Gantmacher's book [100] and Wilkinson [227]. An excellent treatise of nonnegative matrices is in the book by Varga [213] which remains a good reference on iterative methods more three decades after its first publication. Another book with state-of-the-art coverage on iterative methods up to the very beginning of the 1970s is the book by Young [232] which covers M-matrices and related topics in great detail. For a good overview of the linear algebra aspects of matrix theory and a complete proof of Jordan's canonical form, Halmos [117] is recommended. ■