Definition 1.13 A subspace S in $\mathbb{C}^n$ is called *invariant* with respect to a square matrix A if $\mathrm{A}S \subset S$, where $\mathrm{A}S$ is the transformed of S through A. ■

1.8 Similarity Transformations

Definition 1.14 Let C be a square nonsingular matrix having the same order as the matrix A. We say that the matrices A and $\mathrm{C}^{-1}\mathrm{AC}$ are *similar*, and the transformation from A to $\mathrm{C}^{-1}\mathrm{AC}$ is called a *similarity transformation*. Moreover, we say that the two matrices are *unitarily similar* if C is unitary. ■

Two similar matrices share the same spectrum and the same characteristic polynomial. Indeed, it is easy to check that if $(\lambda, \mathbf{x})$ is an eigenvalue-eigenvector pair of A, $(\lambda, \mathrm{C}^{-1}\mathbf{x})$ is the same for the matrix $\mathrm{C}^{-1}\mathrm{AC}$ since

$$(\mathrm{C}^{-1}\mathrm{AC})\mathrm{C}^{-1}\mathbf{x} = \mathrm{C}^{-1}\mathrm{A}\mathbf{x} = \lambda \mathrm{C}^{-1}\mathbf{x}.$$

We notice in particular that the product matrices AB and BA, with $\mathrm{A} \in \mathbb{C}^{n\times m}$ and $\mathrm{B} \in \mathbb{C}^{m\times n}$, are not similar but satisfy the following property (see [Hac94], p.18, Theorem 2.4.6)

$$\sigma(\mathrm{AB})\backslash \{0\} = \sigma(\mathrm{BA})\backslash \{0\},$$

that is, AB and BA share the same spectrum apart from null eigenvalues so that $\rho(\mathrm{AB}) = \rho(\mathrm{BA})$.

The use of similarity transformations aims at reducing the complexity of the problem of evaluating the eigenvalues of a matrix. Indeed, if a given matrix could be transformed into a similar matrix in diagonal or triangular form, the computation of the eigenvalues would be immediate. The main result in this direction is the following theorem (for the proof, see [Dem97], Theorem 4.2).

Property 1.5 (Schur decomposition) *Given* A$\in$ $\mathbb{C}^{n\times n}$, *there exists* U *unitary such that*

$$\mathrm{U}^{-1}\mathrm{AU} = \mathrm{U}^{H}\mathrm{AU} = \begin{bmatrix} \lambda_1 & b_{12} & \ldots & b_{1n} \\ 0 & \lambda_2 & & b_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & \ldots & 0 & \lambda_n \end{bmatrix} = \mathrm{T},$$

where λ_i *are the eigenvalues of* A.

It thus turns out that every matrix A is unitarily similar to an upper triangular matrix. The matrices T and U are not necessarily unique [Hac94]. The Schur decomposition theorem gives rise to several important results; among them, we recall:

1. every hermitian matrix is *unitarily similar* to a diagonal real matrix, that is, when A is hermitian every Schur decomposition of A is diagonal. In such an event, since
$$\mathrm{U}^{-1}\mathrm{AU} = \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n),$$
it turns out that $\mathrm{AU} = \mathrm{U}\Lambda$, that is, $\mathrm{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i$ for $i = 1, \ldots, n$ so that the column vectors of U are the eigenvectors of A. Moreover, since the eigenvectors are orthogonal two by two, it turns out that an hermitian matrix has a system of orthonormal eigenvectors that generates the whole space $\mathbb{C}^n$. Finally, it can be shown that a matrix A of order n is similar to a diagonal matrix D iff the eigenvectors of A form a basis for $\mathbb{C}^n$ [Axe94];
2. a matrix $\mathrm{A} \in \mathbb{C}^{n\times n}$ is normal iff it is unitarily similar to a diagonal matrix. As a consequence, a normal matrix $\mathrm{A} \in \mathbb{C}^{n\times n}$ admits the following *spectral decomposition*: $\mathrm{A} = \mathrm{U}\Lambda\mathrm{U}^H = \sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{u}_i^H$ being U unitary and Λ diagonal [SS90];
3. let A and B be two normal and commutative matrices; then, the generic eigenvalue μ_i of A+B is given by the sum $\lambda_i + \xi_i$, where λ_i and ξ_i are the eigenvalues of A and B associated with the same eigenvector.

There are, of course, nonsymmetric matrices that are similar to diagonal matrices, but these are not unitarily similar (see, e.g., Exercise 7).
The Schur decomposition can be improved as follows (for the proof see, e.g., [Str80], [God66]).

Property 1.6 (Canonical Jordan Form) *Let* A *be any square matrix. Then, there exists a nonsingular matrix* X *which transforms* A *into a block diagonal matrix* J *such that*
$$\mathrm{X}^{-1}\mathrm{AX} = \mathrm{J} = \mathrm{diag}\left(\mathrm{J}_{k_1}(\lambda_1), \mathrm{J}_{k_2}(\lambda_2), \ldots, \mathrm{J}_{k_l}(\lambda_l)\right),$$
which is called canonical Jordan form, λ_j being the eigenvalues of A *and* $\mathrm{J}_k(\lambda) \in \mathbb{C}^{k\times k}$ *a Jordan block of the form* $\mathrm{J}_1(\lambda) = \lambda$ *if* $k = 1$ *and*
$$\mathrm{J}_k(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & \ldots & 0 \\ 0 & \lambda & 1 & \cdots & \vdots \\ \vdots & \ddots & \ddots & 1 & 0 \\ \vdots & & \ddots & \lambda & 1 \\ 0 & \ldots & \ldots & 0 & \lambda \end{bmatrix}, \qquad \textit{for } k > 1.$$

If an eigenvalue is defective, the size of the corresponding Jordan block is greater than one. Therefore, the canonical Jordan form tells us that a matrix can be diagonalized by a similarity transformation iff it is nondefective. For this reason, the nondefective matrices are called *diagonalizable*. In particular, normal matrices are diagonalizable.

Partitioning X by columns, $\mathrm{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, it can be seen that the k_i vectors associated with the Jordan block $\mathrm{J}_{k_i}(\lambda_i)$ satisfy the following recursive relation

$$\begin{aligned} \mathrm{A}\mathbf{x}_l &= \lambda_i \mathbf{x}_l, \qquad\qquad l = \sum_{j=1}^{i-1} m_j + 1, \\ \mathrm{A}\mathbf{x}_j &= \lambda_i \mathbf{x}_j + \mathbf{x}_{j-1},\ j = l+1, \ldots, l-1+k_i, \text{ if } k_i \neq 1. \end{aligned} \tag{1.8}$$

The vectors $\mathbf{x}_i$ are called *principal vectors* or *generalized eigenvectors* of A.

Example 1.6 Let us consider the following matrix

$$\mathrm{A} = \begin{bmatrix} 7/4 & 3/4 & -1/4 & -1/4 & -1/4 & 1/4 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ -1/2 & -1/2 & 5/2 & 1/2 & -1/2 & 1/2 \\ -1/2 & -1/2 & -1/2 & 5/2 & 1/2 & 1/2 \\ -1/4 & -1/4 & -1/4 & -1/4 & 11/4 & 1/4 \\ -3/2 & -1/2 & -1/2 & 1/2 & 1/2 & 7/2 \end{bmatrix}.$$

The Jordan canonical form of A and its associated matrix X are given by

$$\mathrm{J} = \begin{bmatrix} 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}, \qquad \mathrm{X} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Notice that two different Jordan blocks are related to the same eigenvalue ($\lambda = 2$). It is easy to check property (1.8). Consider, for example, the Jordan block associated with the eigenvalue $\lambda_2 = 3$; we have

$$\begin{aligned} \mathrm{A}\mathbf{x}_3 &= [0\ 0\ 3\ 0\ 0\ 3]^T = 3\,[0\ 0\ 1\ 0\ 0\ 1]^T = \lambda_2 \mathbf{x}_3, \\ \mathrm{A}\mathbf{x}_4 &= [0\ 0\ 1\ 3\ 0\ 4]^T = 3\,[0\ 0\ 0\ 1\ 0\ 1]^T + [0\ 0\ 1\ 0\ 0\ 1]^T = \lambda_2 \mathbf{x}_4 + \mathbf{x}_3, \\ \mathrm{A}\mathbf{x}_5 &= [0\ 0\ 0\ 1\ 3\ 4]^T = 3\,[0\ 0\ 0\ 0\ 1\ 1]^T + [0\ 0\ 0\ 1\ 0\ 1]^T = \lambda_2 \mathbf{x}_5 + \mathbf{x}_4. \end{aligned}$$

•

1.9 The Singular Value Decomposition (SVD)

Any matrix can be reduced in diagonal form by a suitable pre and post-multiplication by unitary matrices. Precisely, the following result holds.

Property 1.7 *Let* $\mathrm{A} \in \mathbb{C}^{m\times n}$. *There exist two unitary matrices* $\mathrm{U} \in \mathbb{C}^{m\times m}$ *and* $\mathrm{V} \in \mathbb{C}^{n\times n}$ *such that*

$$\mathrm{U}^H \mathrm{A} \mathrm{V} = \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p) \in \mathbb{R}^{m\times n} \qquad \text{with } p = \min(m, n) \tag{1.9}$$

and $\sigma_1 \geq \ldots \geq \sigma_p \geq 0$. *Formula (1.9) is called Singular Value Decomposition or (SVD) of* A *and the numbers* σ_i *(or* $\sigma_i(\mathrm{A})$*) are called singular values of* A.

If A is a real-valued matrix, U and V will also be real-valued and in (1.9) U^T must be written instead of U^H. The following characterization of the singular values holds

$$\sigma_i(\mathrm{A}) = \sqrt{\lambda_i(\mathrm{A}^H\mathrm{A})}, \quad i = 1, \ldots, p. \tag{1.10}$$

Indeed, from (1.9) it follows that $\mathrm{A} = \mathrm{U}\Sigma\mathrm{V}^H$, $\mathrm{A}^H = \mathrm{V}\Sigma^H\mathrm{U}^H$ so that, U and V being unitary, $\mathrm{A}^H\mathrm{A} = \mathrm{V}\Sigma^H\Sigma\mathrm{V}^H$, that is, $\lambda_i(\mathrm{A}^H\mathrm{A}) = \lambda_i(\Sigma^H\Sigma) = (\sigma_i(\mathrm{A}))^2$. Since AA^H and $\mathrm{A}^H\mathrm{A}$ are hermitian matrices, the columns of U, called the *left singular vectors* of A, turn out to be the eigenvectors of AA^H (see Section 1.8) and, therefore, they are not uniquely defined. The same holds for the columns of V, which are the *right singular vectors* of A.
Relation (1.10) implies that if $\mathrm{A} \in \mathbb{C}^{n\times n}$ is hermitian with eigenvalues given by $\lambda_1, \lambda_2, \ldots, \lambda_n$, then the singular values of A coincide with the modules of the eigenvalues of A. Indeed because $\mathrm{AA}^H = \mathrm{A}^2$, $\sigma_i = \sqrt{\lambda_i^2} = |\lambda_i|$ for $i = 1, \ldots, n$. As far as the rank is concerned, if

$$\sigma_1 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_p = 0,$$

then the rank of A is r, the kernel of A is the span of the column vectors of V, $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$, and the range of A is the span of the column vectors of U, $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$.

Definition 1.15 Suppose that $\mathrm{A} \in \mathbb{C}^{m\times n}$ has rank equal to r and that it admits a SVD of the type $\mathrm{U}^H\mathrm{AV} = \Sigma$. The matrix $\mathrm{A}^\dagger = \mathrm{V}\Sigma^\dagger\mathrm{U}^H$ is called the *Moore-Penrose pseudo-inverse* matrix, being

$$\Sigma^\dagger = \mathrm{diag}\left(\frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_r}, 0, \ldots, 0\right). \tag{1.11}$$

■

The matrix $\mathrm{A}^\dagger$ is also called the *generalized inverse* of A (see Exercise 13). Indeed, if $\mathrm{rank}(\mathrm{A}) = n < m$, then $\mathrm{A}^\dagger = (\mathrm{A}^T\mathrm{A})^{-1}\mathrm{A}^T$, while if $n = m = \mathrm{rank}(\mathrm{A})$, $\mathrm{A}^\dagger = \mathrm{A}^{-1}$. For further properties of $\mathrm{A}^\dagger$, see also Exercise 12.

1.10 Scalar Product and Norms in Vector Spaces

Very often, to quantify errors or measure distances one needs to compute the magnitude of a vector or a matrix. For that purpose we introduce in this section the concept of a vector norm and, in the following one, of a matrix norm. We refer the reader to [Ste73], [SS90] and [Axe94] for the proofs of the properties that are reported hereafter.

Definition 1.16 A *scalar product* on a vector space V defined over K is any map $(\cdot, \cdot)$ acting from $V \times V$ into K which enjoys the following properties:

1. it is linear with respect to the vectors of V, that is

$$(\gamma\mathbf{x} + \lambda\mathbf{z}, \mathbf{y}) = \gamma(\mathbf{x}, \mathbf{y}) + \lambda(\mathbf{z}, \mathbf{y}),\ \forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in V,\ \forall \gamma, \lambda \in K;$$

2. it is *hermitian*, that is, $(\mathbf{y}, \mathbf{x}) = \overline{(\mathbf{x}, \mathbf{y})},\ \forall \mathbf{x}, \mathbf{y} \in V$;
3. it is *positive definite*, that is, $(\mathbf{x}, \mathbf{x}) > 0,\ \forall \mathbf{x} \neq \mathbf{0}$ (in other words, $(\mathbf{x}, \mathbf{x}) \geq 0$, and $(\mathbf{x}, \mathbf{x}) = 0$ if and only if $\mathbf{x} = \mathbf{0}$).

■

In the case $V = \mathbb{C}^n$ (or $\mathbb{R}^n$), an example is provided by the classical Euclidean scalar product given by

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

where $\bar{z}$ denotes the complex conjugate of z.

Moreover, for any given square matrix A of order n and for any $\mathbf{x}$, $\mathbf{y} \in \mathbb{C}^n$ the following relation holds

$$(\mathrm{A}\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathrm{A}^H\mathbf{y}). \tag{1.12}$$

In particular, since for any matrix $\mathrm{Q} \in \mathbb{C}^{n\times n}$, $(\mathrm{Q}\mathbf{x}, \mathrm{Q}\mathbf{y}) = (\mathbf{x}, \mathrm{Q}^H\mathrm{Q}\mathbf{y})$, one gets

Property 1.8 *Unitary matrices preserve the Euclidean scalar product, that is,* $(\mathrm{Q}\mathbf{x}, \mathrm{Q}\mathbf{y}) = (\mathbf{x}, \mathbf{y})$ *for any unitary matrix* Q *and for any pair of vectors* $\mathbf{x}$ *and* $\mathbf{y}$.

Definition 1.17 Let V be a vector space over K. We say that the map $\|\cdot\|$ from V into $\mathbb{R}$ is a *norm* on V if the following axioms are satisfied:

1. (i) $\|\mathbf{v}\| \geq 0\ \forall \mathbf{v} \in V$ and (ii) $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$;
2. $\|\alpha\mathbf{v}\| = |\alpha|\|\mathbf{v}\|\ \ \forall \alpha \in K, \forall \mathbf{v} \in V$ (homogeneity property);
3. $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|\ \ \forall \mathbf{v}, \mathbf{w} \in V$ (triangular inequality),

where $|\alpha|$ denotes the absolute value of α if $K = \mathbb{R}$, the module of α if $K = \mathbb{C}$. ■

The pair $(V, \|\cdot\|)$ is called a *normed space*. We shall distinguish among norms by a suitable subscript at the margin of the double bar symbol. In the case the map $|\cdot|$ from V into $\mathbb{R}$ enjoys only the properties $1(i)$, 2 and 3 we shall call such a map a *seminorm*. Finally, we shall call a *unit vector* any vector of V having unit norm.

An example of a normed space is $\mathbb{R}^n$, equipped for instance by the *p-norm* (or *Hölder norm*); this latter is defined for a vector $\mathbf{x}$ of components $\{x_i\}$ as

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}, \qquad \text{for } 1 \leq p < \infty. \tag{1.13}$$

Notice that the limit as p goes to infinity of $\|\mathbf{x}\|_p$ exists, is finite, and equals the maximum module of the components of $\mathbf{x}$. Such a limit defines in turn a norm, called the *infinity norm* (or *maximum norm*), given by

$$\|\mathbf{x}\|_\infty = \max_{1\leq i\leq n} |x_i|.$$

When $p = 2$, from (1.13) the standard definition of *Euclidean norm* is recovered

$$\|\mathbf{x}\|_2 = (\mathbf{x},\mathbf{x})^{1/2} = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2} = \left(\mathbf{x}^T\mathbf{x}\right)^{1/2},$$

for which the following property holds.

Property 1.9 (Cauchy-Schwarz inequality) *For any pair* $\mathbf{x},\mathbf{y} \in \mathbb{R}^n$*,*

$$|(\mathbf{x},\mathbf{y})| = |\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|_2 \ \|\mathbf{y}\|_2, \tag{1.14}$$

where strict equality holds iff $\mathbf{y} = \alpha\mathbf{x}$ *for some* $\alpha \in \mathbb{R}$*.*

We recall that the scalar product in $\mathbb{R}^n$ can be related to the p-norms introduced over $\mathbb{R}^n$ in (1.13) by the *Hölder inequality*

$$|(\mathbf{x},\mathbf{y})| \leq \|\mathbf{x}\|_p\|\mathbf{y}\|_q, \ \text{ with } \frac{1}{p} + \frac{1}{q} = 1.$$

In the case where V is a finite-dimensional space the following property holds (for a sketch of the proof, see Exercise 14).

Property 1.10 *Any vector norm* $\|\cdot\|$ *defined on* V *is a continuous function of its argument, namely,* $\forall\varepsilon > 0$, $\exists C > 0$ *such that if* $\|\mathbf{x} - \widehat{\mathbf{x}}\| \leq \varepsilon$ *then* $|\ \|\mathbf{x}\| - \|\widehat{\mathbf{x}}\|\ | \leq C\varepsilon$*, for any* $\mathbf{x}$, $\widehat{\mathbf{x}} \in V$*.*

New norms can be easily built using the following result.

Property 1.11 *Let* $\|\cdot\|$ *be a norm of* $\mathbb{R}^n$ *and* $\mathrm{A} \in \mathbb{R}^{n\times n}$ *be a matrix with* n *linearly independent columns. Then, the function* $\|\cdot\|_{\mathrm{A}^2}$ *acting from* $\mathbb{R}^n$ *into* $\mathbb{R}$ *defined as*

$$\|\mathbf{x}\|_{\mathrm{A}^2} = \|\mathrm{A}\mathbf{x}\| \qquad \forall\mathbf{x} \in \mathbb{R}^n,$$

is a norm of $\mathbb{R}^n$*.*

Two vectors $\mathbf{x}$, $\mathbf{y}$ in V are said to be *orthogonal* if $(\mathbf{x},\mathbf{y}) = 0$. This statement has an immediate geometric interpretation when $V = \mathbb{R}^2$ since in such a case

$$(\mathbf{x},\mathbf{y}) = \|\mathbf{x}\|_2\|\mathbf{y}\|_2\cos(\vartheta),$$

Table 1.1. Equivalence constants for the main norms of $\mathbb{R}^n$

c_{pq}	$q=1$	$q=2$	$q=\infty$
$p=1$	1	1	1
$p=2$	$n^{-1/2}$	1	1
$p=\infty$	n^{-1}	$n^{-1/2}$	1

C_{pq}	$q=1$	$q=2$	$q=\infty$
$p=1$	1	$n^{1/2}$	n
$p=2$	1	1	$n^{1/2}$
$p=\infty$	1	1	1

where ϑ is the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$. As a consequence, if $(\mathbf{x},\mathbf{y}) = 0$ then ϑ is a right angle and the two vectors are orthogonal in the geometric sense.

Definition 1.18 Two norms $\|\cdot\|_p$ and $\|\cdot\|_q$ on V are *equivalent* if there exist two positive constants c_{pq} and C_{pq} such that

$$c_{pq}\|\mathbf{x}\|_q \leq \|\mathbf{x}\|_p \leq C_{pq}\|\mathbf{x}\|_q \ \forall \mathbf{x} \in V.$$

■

In a finite-dimensional normed space all norms are equivalent. In particular, if $V = \mathbb{R}^n$ it can be shown that for the p-norms, with $p = 1$, 2, and ∞, the constants c_{pq} and C_{pq} take the value reported in Table 1.1.

In this book we shall often deal with sequences of vectors and with their *convergence*. For this purpose, we recall that a sequence of vectors $\left\{\mathbf{x}^{(k)}\right\}$ in a vector space V having finite dimension n, converges to a vector $\mathbf{x}$, and we write $\lim\limits_{k\to\infty} \mathbf{x}^{(k)} = \mathbf{x}$ if

$$\lim_{k\to\infty} x_i^{(k)} = x_i,\ i = 1,\ldots,n, \tag{1.15}$$

where $x_i^{(k)}$ and x_i are the components of the corresponding vectors with respect to a basis of V. If $V = \mathbb{R}^n$, due to the uniqueness of the limit of a sequence of real numbers, (1.15) implies also the uniqueness of the limit, if existing, of a sequence of vectors.

We further notice that in a finite-dimensional space all the norms are topologically equivalent in the sense of convergence, namely, given a sequence of vectors $\mathbf{x}^{(k)}$, we have that

$$|||\mathbf{x}^{(k)}||| \to 0 \ \Leftrightarrow \|\mathbf{x}^{(k)}\| \to 0 \text{ if } k \to \infty,$$

where $|||\cdot|||$ and $\|\cdot\|$ are any two vector norms. As a consequence, we can establish the following link between norms and limits.

Property 1.12 *Let $\|\cdot\|$ be a norm in a finite dimensional space V. Then*

$$\lim_{k\to\infty} \mathbf{x}^{(k)} = \mathbf{x} \Leftrightarrow \lim_{k\to\infty} \|\mathbf{x} - \mathbf{x}^{(k)}\| = 0,$$

where $\mathbf{x} \in V$ and $\left\{\mathbf{x}^{(k)}\right\}$ is a sequence of elements of V.

1.11 Matrix Norms

Definition 1.19 A *matrix norm* is a mapping $\|\cdot\| : \mathbb{R}^{m\times n} \to \mathbb{R}$ such that:

1. $\|\mathrm{A}\| \geq 0\ \forall \mathrm{A} \in \mathbb{R}^{m\times n}$ and $\|\mathrm{A}\| = 0$ if and only if $\mathrm{A} = 0$;
2. $\|\alpha \mathrm{A}\| = |\alpha| \|\mathrm{A}\|\ \ \forall \alpha \in \mathbb{R}, \forall \mathrm{A} \in \mathbb{R}^{m\times n}$ (homogeneity);
3. $\|\mathrm{A} + \mathrm{B}\| \leq \|\mathrm{A}\| + \|\mathrm{B}\|\ \ \forall \mathrm{A}, \mathrm{B} \in \mathbb{R}^{m\times n}$ (triangular inequality).

■

Unless otherwise specified we shall employ the same symbol $\|\cdot\|$, to denote matrix norms and vector norms.

We can better characterize the matrix norms by introducing the concepts of compatible norm and norm induced by a vector norm.

Definition 1.20 We say that a matrix norm $\|\cdot\|$ is *compatible* or *consistent* with a vector norm $\|\cdot\|$ if

$$\|\mathrm{A}\mathbf{x}\| \leq \|\mathrm{A}\|\ \|\mathbf{x}\|, \qquad \forall \mathbf{x} \in \mathbb{R}^n. \tag{1.16}$$

More generally, given three norms, all denoted by $\|\cdot\|$, albeit defined on $\mathbb{R}^m$, $\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$, respectively, we say that they are consistent if $\forall \mathbf{x} \in \mathbb{R}^n$, $\mathrm{A}\mathbf{x} = \mathbf{y} \in \mathbb{R}^m$, $\mathrm{A} \in \mathbb{R}^{m\times n}$, we have that $\|\mathbf{y}\| \leq \|\mathrm{A}\|\ \|\mathbf{x}\|$. ■

In order to single out matrix norms of practical interest, the following property is in general required

Definition 1.21 We say that a matrix norm $\|\cdot\|$ is *sub-multiplicative* if $\forall \mathrm{A} \in \mathbb{R}^{n\times m}$, $\forall \mathrm{B} \in \mathbb{R}^{m\times q}$

$$\|\mathrm{AB}\| \leq \|\mathrm{A}\|\ \|\mathrm{B}\|. \tag{1.17}$$

■

This property is not satisfied by any matrix norm. For example (taken from [GL89]), the norm $\|\mathrm{A}\|_\Delta = \max |a_{ij}|$ for $i = 1, \ldots, n$, $j = 1, \ldots, m$ does not satisfy (1.17) if applied to the matrices

$$\mathrm{A} = \mathrm{B} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

since $2 = \|\mathrm{AB}\|_\Delta > \|\mathrm{A}\|_\Delta \|\mathrm{B}\|_\Delta = 1$.
Notice that, given a certain sub-multiplicative matrix norm $\|\cdot\|_\alpha$, there always exists a consistent vector norm. For instance, given any fixed vector $\mathbf{y} \neq \mathbf{0}$ in $\mathbb{C}^n$, it suffices to define the consistent vector norm as

$$\|\mathbf{x}\| = \|\mathbf{x}\mathbf{y}^H\|_\alpha \qquad \mathbf{x} \in \mathbb{C}^n.$$

As a consequence, in the case of sub-multiplicative matrix norms it is no longer necessary to explicitly specify the vector norm with respect to the matrix norm is consistent.

Example 1.7 The norm

$$\|A\|_F = \sqrt{\sum_{i,j=1}^{n} |a_{ij}|^2} = \sqrt{\mathrm{tr}(AA^H)} \tag{1.18}$$

is a matrix norm called the *Frobenius norm* (or *Euclidean norm* in $\mathbb{C}^{n^2}$) and is compatible with the Euclidean vector norm $\|\cdot\|_2$. Indeed,

$$\|A\mathbf{x}\|_2^2 = \sum_{i=1}^{n} \left| \sum_{j=1}^{n} a_{ij} x_j \right|^2 \leq \sum_{i=1}^{n} \left(\sum_{j=1}^{n} |a_{ij}|^2 \sum_{j=1}^{n} |x_j|^2 \right) = \|A\|_F^2 \|\mathbf{x}\|_2^2.$$

Notice that for such a norm $\|I_n\|_F = \sqrt{n}$. •

In view of the definition of a natural norm, we recall the following theorem.

Theorem 1.1 *Let $\|\cdot\|$ be a vector norm. The function*

$$\|A\| = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} \tag{1.19}$$

is a matrix norm called induced matrix norm or natural matrix norm.

Proof. We start by noticing that (1.19) is equivalent to

$$\|A\| = \sup_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|. \tag{1.20}$$

Indeed, one can define for any $\mathbf{x} \neq \mathbf{0}$ the unit vector $\mathbf{u} = \mathbf{x}/\|\mathbf{x}\|$, so that (1.19) becomes

$$\|A\| = \sup_{\|\mathbf{u}\|=1} \|A\mathbf{u}\| = \|A\mathbf{w}\| \qquad \text{with } \|\mathbf{w}\| = 1.$$

This being taken as given, let us check that (1.19) (or, equivalently, (1.20)) is actually a norm, making direct use of Definition 1.19.

1. If $\|A\mathbf{x}\| \geq 0$, then it follows that $\|A\| = \sup_{\|\mathbf{x}\|=1} \|A\mathbf{x}\| \geq 0$. Moreover

$$\|A\| = \sup_{\mathbf{x} \neq \mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} = 0 \Leftrightarrow \|A\mathbf{x}\| = 0 \ \forall \mathbf{x} \neq \mathbf{0},$$

   and $A\mathbf{x} = \mathbf{0} \ \forall \mathbf{x} \neq \mathbf{0}$ if and only if A=0; therefore $\|A\| = 0 \Leftrightarrow A = 0$.
2. Given a scalar α,

$$\|\alpha A\| = \sup_{\|\mathbf{x}\|=1} \|\alpha A\mathbf{x}\| = |\alpha| \sup_{\|\mathbf{x}\|=1} \|A\mathbf{x}\| = |\alpha| \ \|A\|.$$

3. Finally, triangular inequality holds. Indeed, by definition of supremum, if $\mathbf{x} \neq \mathbf{0}$ then

$$\frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} \leq \|A\| \Rightarrow \|A\mathbf{x}\| \leq \|A\|\|\mathbf{x}\|,$$

   so that, taking $\mathbf{x}$ with unit norm, one gets

$$\|(A+B)\mathbf{x}\| \leq \|A\mathbf{x}\| + \|B\mathbf{x}\| \leq \|A\| + \|B\|,$$

   from which it follows that $\|A+B\| = \sup_{\|\mathbf{x}\|=1} \|(A+B)\mathbf{x}\| \leq \|A\| + \|B\|$.

◇

Relevant instances of induced matrix norms are the so-called *p-norms* defined as

$$\|A\|_p = \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p}.$$

The 1-norm and the infinity norm are easily computable since

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

and they are called the *column sum norm* and the *row sum norm*, respectively.

Moreover, we have $\|A\|_1 = \|A^T\|_\infty$ and, if A is self-adjoint or real symmetric, $\|A\|_1 = \|A\|_\infty$.

A special discussion is deserved by the *2-norm* or *spectral norm* for which the following theorem holds.

Theorem 1.2 *Let* $\sigma_1(A)$ *be the largest singular value of* A. *Then*

$$\|A\|_2 = \sqrt{\rho(A^H A)} = \sqrt{\rho(AA^H)} = \sigma_1(A). \qquad (1.21)$$

In particular, if A *is hermitian (or real and symmetric), then*

$$\|A\|_2 = \rho(A), \qquad (1.22)$$

while, if A *is unitary,* $\|A\|_2 = 1$.

Proof. Since $A^H A$ is hermitian, there exists a unitary matrix U such that

$$U^H A^H A U = \mathrm{diag}(\mu_1, \ldots, \mu_n),$$

where μ_i are the (positive) eigenvalues of $A^H A$. Let $\mathbf{y} = U^H \mathbf{x}$, then

$$\begin{aligned} \|A\|_2 &= \sup_{\mathbf{x}\neq\mathbf{0}} \sqrt{\frac{(A^H A\mathbf{x}, \mathbf{x})}{(\mathbf{x}, \mathbf{x})}} = \sup_{\mathbf{y}\neq\mathbf{0}} \sqrt{\frac{(U^H A^H A U\mathbf{y}, \mathbf{y})}{(\mathbf{y}, \mathbf{y})}} \\ &= \sup_{\mathbf{y}\neq\mathbf{0}} \sqrt{\sum_{i=1}^{n} \mu_i |y_i|^2 \Big/ \sum_{i=1}^{n} |y_i|^2} = \sqrt{\max_{i=1,\ldots,n} |\mu_i|}, \end{aligned}$$

from which (1.21) follows, thanks to (1.10).

If A is hermitian, the same considerations as above apply directly to A.

Finally, if A is unitary, we have

$$\|A\mathbf{x}\|_2^2 = (A\mathbf{x}, A\mathbf{x}) = (\mathbf{x}, A^H A\mathbf{x}) = \|\mathbf{x}\|_2^2,$$

so that $\|A\|_2 = 1$. ◇

As a consequence, the computation of $\|A\|_2$ is much more expensive than that of $\|A\|_\infty$ or $\|A\|_1$. However, if only an estimate of $\|A\|_2$ is required, the following relations can be profitably employed in the case of square matrices

$$\max_{i,j}|a_{ij}| \leq \|\mathrm{A}\|_2 \leq n \max_{i,j}|a_{ij}|,$$
$$\frac{1}{\sqrt{n}}\|\mathrm{A}\|_\infty \leq \|\mathrm{A}\|_2 \leq \sqrt{n}\|\mathrm{A}\|_\infty,$$
$$\frac{1}{\sqrt{n}}\|\mathrm{A}\|_1 \leq \|\mathrm{A}\|_2 \leq \sqrt{n}\|\mathrm{A}\|_1,$$
$$\|\mathrm{A}\|_2 \leq \sqrt{\|\mathrm{A}\|_1 \ \|\mathrm{A}\|_\infty}.$$

For other estimates of similar type we refer to Exercise 17. Moreover, if A is normal then $\|\mathrm{A}\|_2 \leq \|\mathrm{A}\|_p$ for any n and all $p \geq 2$.

Theorem 1.3 *Let $|||\cdot|||$ be a matrix norm induced by a vector norm $\|\cdot\|$. Then, the following relations hold:*

1. $\|\mathrm{A}\mathbf{x}\| \leq |||\mathrm{A}||| \ \|\mathbf{x}\|$, *that is,* $|||\cdot|||$ *is a norm compatible with* $\|\cdot\|$;
2. $|||\mathrm{I}||| = 1$;
3. $|||\mathrm{AB}||| \leq |||\mathrm{A}||| \ |||\mathrm{B}|||$, *that is,* $|||\cdot|||$ *is sub-multiplicative.*

Proof. Part 1 of the theorem is already contained in the proof of Theorem 1.1, while part 2 follows from the fact that $|||\mathrm{I}||| = \sup_{\mathbf{x}\neq\mathbf{0}}\|\mathrm{I}\mathbf{x}\|/\|\mathbf{x}\| = 1$. Part 3 is simple to check. ◇

Notice that the p-norms are sub-multiplicative. Moreover, we remark that the sub-multiplicativity property by itself would only allow us to conclude that $|||\mathrm{I}||| \geq 1$. Indeed, $|||\mathrm{I}||| = |||\mathrm{I}\cdot\mathrm{I}||| \leq |||\mathrm{I}|||^2$.

1.11.1 Relation between Norms and the Spectral Radius of a Matrix

We next recall some results that relate the spectral radius of a matrix to matrix norms and that will be widely employed in Chapter 4.

Theorem 1.4 *Let $\|\cdot\|$ be a consistent matrix norm; then*

$$\rho(\mathrm{A}) \leq \|\mathrm{A}\| \qquad \forall \mathrm{A} \in \mathbb{C}^{n\times n}.$$

Proof. Let λ be an eigenvalue of A and $\mathbf{v} \neq \mathbf{0}$ an associated eigenvector. As a consequence, since $\|\cdot\|$ is consistent, we have

$$|\lambda| \ \|\mathbf{v}\| = \|\lambda\mathbf{v}\| = \|\mathrm{A}\mathbf{v}\| \leq \|\mathrm{A}\| \ \|\mathbf{v}\|,$$

so that $|\lambda| \leq \|\mathrm{A}\|$. ◇

More precisely, the following property holds (see for the proof [IK66], p. 12, Theorem 3).

Property 1.13 *Let $\mathrm{A} \in \mathbb{C}^{n\times n}$ and $\varepsilon > 0$. Then, there exists an induced matrix norm $\|\cdot\|_{\mathrm{A},\varepsilon}$ (depending on ε) such that*

$$\|\mathrm{A}\|_{\mathrm{A},\varepsilon} \leq \rho(\mathrm{A}) + \varepsilon.$$

As a result, having fixed an arbitrarily small tolerance, there always exists a matrix norm which is arbitrarily close to the spectral radius of A, namely

$$\rho(\mathrm{A}) = \inf_{\|\cdot\|} \|\mathrm{A}\|, \tag{1.23}$$

the infimum being taken on the set of all the consistent norms.

For the sake of clarity, we notice that the spectral radius is a sub-multiplicative seminorm, since it is not true that $\rho(\mathrm{A}) = 0$ iff $\mathrm{A} = 0$. As an example, any triangular matrix with null diagonal entries clearly has spectral radius equal to zero. Moreover, we have the following result.

Property 1.14 *Let* A *be a square matrix and let* $\|\cdot\|$ *be a consistent norm. Then*

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

1.11.2 Sequences and Series of Matrices

A sequence of matrices $\left\{\mathrm{A}^{(k)}\right\} \in \mathbb{R}^{n\times n}$ is said to *converge* to a matrix $\mathrm{A} \in \mathbb{R}^{n\times n}$ if

$$\lim_{k\to\infty} \|\mathrm{A}^{(k)} - \mathrm{A}\| = 0.$$

The choice of the norm does not influence the result since in $\mathbb{R}^{n\times n}$ all norms are equivalent. In particular, when studying the convergence of iterative methods for solving linear systems (see Chapter 4), one is interested in the so-called *convergent matrices* for which

$$\lim_{k\to\infty} \mathrm{A}^k = 0,$$

0 being the null matrix. The following theorem holds.

Theorem 1.5 *Let* A *be a square matrix; then*

$$\lim_{k\to\infty} \mathrm{A}^k = 0 \Leftrightarrow \rho(\mathrm{A}) < 1. \tag{1.24}$$

Moreover, the geometric series $\sum_{k=0}^{\infty} \mathrm{A}^k$ *is convergent iff* $\rho(\mathrm{A}) < 1$. *In such a case*

$$\sum_{k=0}^{\infty} \mathrm{A}^k = (\mathrm{I} - \mathrm{A})^{-1}. \tag{1.25}$$

As a result, if $\rho(\mathrm{A}) < 1$ *the matrix* $\mathrm{I} - \mathrm{A}$ *is invertible and the following inequalities hold*

$$\frac{1}{1 + \|\mathrm{A}\|} \leq \|(\mathrm{I} - \mathrm{A})^{-1}\| \leq \frac{1}{1 - \|\mathrm{A}\|}, \tag{1.26}$$

where $\|\cdot\|$ *is an induced matrix norm such that* $\|\mathrm{A}\| < 1$.

Proof. Let us prove (1.24). Let $\rho(\mathrm{A}) < 1$, then $\exists \varepsilon > 0$ such that $\rho(\mathrm{A}) < 1-\varepsilon$ and thus, thanks to Property 1.13, there exists an induced matrix norm $\|\cdot\|$ such that $\|\mathrm{A}\| \leq \rho(\mathrm{A})+\varepsilon < 1$. From the fact that $\|\mathrm{A}^k\| \leq \|\mathrm{A}\|^k < 1$ and from the definition of convergence it turns out that as $k \to \infty$ the sequence $\left\{\mathrm{A}^k\right\}$ tends to zero. Conversely, assume that $\lim_{k\to\infty} \mathrm{A}^k = 0$ and let λ denote an eigenvalue of A. Then, $\mathrm{A}^k\mathbf{x} = \lambda^k\mathbf{x}$, being $\mathbf{x}(\neq\mathbf{0})$ an eigenvector associated with λ, so that $\lim_{k\to\infty} \lambda^k = 0$. As a consequence, $|\lambda| < 1$ and because this is true for a generic eigenvalue one gets $\rho(\mathrm{A}) < 1$ as desired. Relation (1.25) can be obtained noting first that the eigenvalues of I$-$A are given by $1-\lambda(\mathrm{A})$, $\lambda(\mathrm{A})$ being the generic eigenvalue of A. On the other hand, since $\rho(\mathrm{A}) < 1$, we deduce that I$-$A is nonsingular. Then, from the identity

$$(\mathrm{I}-\mathrm{A})(\mathrm{I}+\mathrm{A}+\ldots+\mathrm{A}^n) = (\mathrm{I}-\mathrm{A}^{n+1})$$

and taking the limit for n tending to infinity the thesis follows since

$$(\mathrm{I}-\mathrm{A})\sum_{k=0}^{\infty}\mathrm{A}^k = \mathrm{I}.$$

Finally, thanks to Theorem 1.3, the equality $\|\mathrm{I}\| = 1$ holds, so that

$$1 = \|\mathrm{I}\| \leq \|\mathrm{I}-\mathrm{A}\| \; \|(\mathrm{I}-\mathrm{A})^{-1}\| \leq (1+\|\mathrm{A}\|)\, \|(\mathrm{I}-\mathrm{A})^{-1}\|,$$

giving the first inequality in (1.26). As for the second part, noting that $\mathrm{I} = \mathrm{I}-\mathrm{A}+\mathrm{A}$ and multiplying both sides on the right by $(\mathrm{I}-\mathrm{A})^{-1}$, one gets $(\mathrm{I}-\mathrm{A})^{-1} = \mathrm{I} + \mathrm{A}(\mathrm{I}-\mathrm{A})^{-1}$. Passing to the norms, we obtain

$$\|(\mathrm{I}-\mathrm{A})^{-1}\| \leq 1 + \|\mathrm{A}\| \; \|(\mathrm{I}-\mathrm{A})^{-1}\|,$$

and thus the second inequality, since $\|\mathrm{A}\| < 1$. ◇

Remark 1.1 The assumption that there exists an induced matrix norm such that $\|\mathrm{A}\| < 1$ is justified by Property 1.13, recalling that A is convergent and, therefore, $\rho(\mathrm{A}) < 1$. ■

Notice that (1.25) suggests an algorithm to approximate the inverse of a matrix by a truncated series expansion.

1.12 Positive Definite, Diagonally Dominant and M-matrices

Definition 1.22 A matrix $\mathrm{A} \in \mathbb{C}^{n\times n}$ is *positive definite in* $\mathbb{C}^n$ if the number $(\mathrm{A}\mathbf{x},\mathbf{x})$ is real and positive $\forall\mathbf{x} \in \mathbb{C}^n$, $\mathbf{x} \neq \mathbf{0}$. A matrix $\mathrm{A} \in \mathbb{R}^{n\times n}$ is *positive definite in* $\mathbb{R}^n$ if $(\mathrm{A}\mathbf{x},\mathbf{x}) > 0$ $\forall\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} \neq \mathbf{0}$. If the strict inequality is substituted by the weak one ($\geq$) the matrix is called *positive semi-definite.* ■

Example 1.8 Matrices that are positive definite in $\mathbb{R}^n$ are not necessarily symmetric. An instance is provided by matrices of the form

$$\mathrm{A} = \begin{bmatrix} 2 & \alpha \\ -2-\alpha & 2 \end{bmatrix} \tag{1.27}$$

for $\alpha \neq -1$. Indeed, for any nonnull vector $\mathbf{x} = (x_1, x_2)^T$ in $\mathbb{R}^2$

$$(\mathrm{A}\mathbf{x}, \mathbf{x}) = 2(x_1^2 + x_2^2 - x_1 x_2) > 0.$$

Notice that A is *not* positive definite in $\mathbb{C}^2$. Indeed, if we take a complex vector $\mathbf{x}$ we find out that the number $(\mathrm{A}\mathbf{x}, \mathbf{x})$ is not real-valued in general. •

Definition 1.23 Let $\mathrm{A} \in \mathbb{R}^{n\times n}$. The matrices

$$\mathrm{A}_S = \frac{1}{2}(\mathrm{A} + \mathrm{A}^T), \ \mathrm{A}_{SS} = \frac{1}{2}(\mathrm{A} - \mathrm{A}^T)$$

are respectively called the *symmetric part* and the *skew-symmetric part* of A. Obviously, $\mathrm{A} = \mathrm{A}_S + \mathrm{A}_{SS}$. If $\mathrm{A} \in \mathbb{C}^{n\times n}$, the definitions modify as follows: $\mathrm{A}_S = \frac{1}{2}(\mathrm{A} + \mathrm{A}^H)$ and $\mathrm{A}_{SS} = \frac{1}{2}(\mathrm{A} - \mathrm{A}^H)$. ■

The following property holds

Property 1.15 *A real matrix* A *of order* n *is positive definite iff its symmetric part* A_S *is positive definite.*

Indeed, it suffices to notice that, due to (1.12) and the definition of A_{SS}, $\mathbf{x}^T \mathrm{A}_{SS}\mathbf{x} = 0 \ \forall \mathbf{x} \in \mathbb{R}^n$. For instance, the matrix in (1.27) has a positive definite symmetric part, since

$$\mathrm{A}_S = \frac{1}{2}(\mathrm{A} + \mathrm{A}^T) = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

This holds more generally (for the proof see [Axe94]).

Property 1.16 *Let* $\mathrm{A} \in \mathbb{C}^{n\times n}$ *(respectively,* $\mathrm{A} \in \mathbb{R}^{n\times n}$*); if* $(\mathrm{A}\mathbf{x}, \mathbf{x})$ *is real-valued* $\forall \mathbf{x} \in \mathbb{C}^n$*, then* A *is hermitian (respectively, symmetric).*

An immediate consequence of the above results is that matrices that are positive definite in $\mathbb{C}^n$ do satisfy the following characterizing property.

Property 1.17 *A square matrix* A *of order* n *is positive definite in* $\mathbb{C}^n$ *iff it is hermitian and has positive eigenvalues. Thus, a positive definite matrix is nonsingular.*

In the case of positive definite real matrices in $\mathbb{R}^n$, results more specific than those presented so far hold only if the matrix is also symmetric (this is the reason why many textbooks deal only with symmetric positive definite matrices). In particular

Property 1.18 *Let* $\mathrm{A} \in \mathbb{R}^{n\times n}$ *be symmetric. Then,* A *is positive definite iff one of the following properties is satisfied:*

1. $(\mathrm{A}\mathbf{x},\mathbf{x}) > 0\ \forall \mathbf{x} \neq \mathbf{0}$ *with* $\mathbf{x}\in \mathbb{R}^n$*;*
2. *the eigenvalues of the principal submatrices of* A *are all positive;*
3. *the dominant principal minors of* A *are all positive (Sylvester criterion);*
4. *there exists a nonsingular matrix* H *such that* $\mathrm{A} = \mathrm{H}^T\mathrm{H}$*.*

All the diagonal entries of a positive definite matrix are positive. Indeed, if $\mathbf{e}_i$ is the i-th vector of the canonical basis of $\mathbb{R}^n$, then $\mathbf{e}_i^T\mathrm{A}\mathbf{e}_i = a_{ii} > 0$.

Moreover, it can be shown that if A is symmetric positive definite, the entry with the largest module must be a diagonal entry (these last two properties are therefore necessary conditions for a matrix to be positive definite).

We finally notice that if A is symmetric positive definite and $\mathrm{A}^{1/2}$ is the only positive definite matrix that is a solution of the matrix equation $\mathrm{X}^2 = \mathrm{A}$, the norm

$$\|\mathbf{x}\|_{\mathrm{A}} = \|\mathrm{A}^{1/2}\mathbf{x}\|_2 = (\mathrm{A}\mathbf{x},\mathbf{x})^{1/2} \tag{1.28}$$

defines a vector norm, called the *energy norm* of the vector $\mathbf{x}$. Related to the energy norm is the *energy scalar product* given by $(\mathbf{x},\mathbf{y})_{\mathrm{A}} = (\mathrm{A}\mathbf{x},\mathbf{y})$.

Definition 1.24 A matrix $\mathrm{A}\in \mathbb{R}^{n\times n}$ is called *diagonally dominant by rows* if

$$|a_{ii}| \geq \sum_{j=1,j\neq i}^{n} |a_{ij}|, \text{ with } i = 1,\dots,n,$$

while it is called *diagonally dominant by columns* if

$$|a_{ii}| \geq \sum_{j=1,j\neq i}^{n} |a_{ji}|, \text{ with } i = 1,\dots,n.$$

If the inequalities above hold in a strict sense, A is called *strictly diagonally dominant* (by rows or by columns, respectively). ■

A strictly diagonally dominant matrix that is symmetric with positive diagonal entries is also positive definite.

Definition 1.25 A nonsingular matrix $\mathrm{A} \in \mathbb{R}^{n\times n}$ is an *M-matrix* if $a_{ij} \leq 0$ for $i \neq j$ and if all the entries of its inverse are nonnegative. ■

M-matrices enjoy the so-called *discrete maximum principle*, that is, if A is an M-matrix and $\mathrm{A}\mathbf{x} \leq \mathbf{0}$, then $\mathbf{x} \leq \mathbf{0}$ (where the inequalities are meant componentwise). In this connection, the following result can be useful.

Property 1.19 (M-criterion) *Let a matrix* A *satisfy* $a_{ij} \leq 0$ *for* $i \neq j$*. Then* A *is an M-matrix if and only if there exists a vector* $\mathbf{w} > 0$ *such that* $\mathrm{A}\mathbf{w} > 0$*.*