1 Matrix Theory

We collect in this chapter useful background material on matrix theory and linear algebra. The emphasis is on results that are needed for future developments. Among other concepts, we review symmetric and non-negative definite matrices, range spaces and nullspaces, as well as several matrix decompositions including the spectral decomposition, the triangular decomposition, the QR decomposition, and the singular value decomposition (SVD). We also discuss vector and matrix norms, Kronecker products, Schur complements, and the useful Rayleigh-Ritz characterization of the eigenvalues of symmetric matrices.

1.1 SYMMETRIC MATRICES

Symmetric and non-negative definite matrices play a prominent role in data analysis. We review some of their properties in this section. Thus, consider an arbitrary square matrix of size $N \times N$ with real entries, written as $A \in \mathbb{R}^{N\times N}$. The transpose of A is denoted by A^{T} and is obtained by transforming the rows of A into columns of A^{T}. For example,

$$A = \begin{bmatrix} 1 & -1 & 3 \\ -2 & 4 & 5 \\ 0 & 6 & 8 \end{bmatrix} \quad \Longrightarrow \quad A^{\mathsf{T}} = \begin{bmatrix} 1 & -2 & 0 \\ -1 & 4 & 6 \\ 3 & 5 & 8 \end{bmatrix} \tag{1.1}$$

The matrix A is said to be symmetric if it happens to coincide with its matrix transpose, i.e., if it satisfies

$$A = A^{\mathsf{T}}, \qquad (\textbf{symmetry}) \tag{1.2}$$

Real eigenvalues

A useful property of symmetric matrices is that they can only have *real* eigenvalues. To see this, let u represent a column eigenvector of A corresponding to some eigenvalue λ, i.e., u is nonzero and satisfies along with λ the relation:

$$Au = \lambda u \tag{1.3}$$

The eigenvector u may be complex-valued so that, in general, $u \in \mathbb{C}^N$. Let the symbol $*$ denote the operation of complex conjugate transposition, so that u^*

is the row vector that is obtained by transposing u and replacing its entries by their complex conjugate values, e.g.,

$$u \stackrel{\Delta}{=} \begin{bmatrix} 1+j \\ 2 \\ -2+3j \end{bmatrix} \implies u^* = \begin{bmatrix} 1-j & 2 & -2-3j \end{bmatrix} \tag{1.4}$$

where $j \stackrel{\Delta}{=} \sqrt{-1}$. The same complex conjugation operation can be applied to matrices as well so that, for example,

$$B = \begin{bmatrix} 1 & j & -2+j \\ 3-j & 1-2j & 0 \end{bmatrix} \implies B^* = \begin{bmatrix} 1 & 3+j \\ -j & 1+2j \\ -2-j & 0 \end{bmatrix} \tag{1.5}$$

Returning to (1.3) and multiplying from the left by the row vector u^* we get

$$u^*Au = \lambda u^*u \;=\; \lambda\|u\|^2 \tag{1.6}$$

where the notation $\|\cdot\|$ denotes the Euclidean norm of its vector argument. Note that the quantity u^*Au is a scalar. Moreover, it is real-valued because it coincides with its complex conjugate value:

$$(u^*Au)^* \;=\; u^*A^*(u^*)^* \;=\; u^*Au \tag{1.7}$$

where in the last step we used the fact that $A^* = A$ since A is real-valued and symmetric. Therefore, u^*Au is real and, from equality (1.6), we conclude that $\lambda\|u\|^2$ must also be real. But since $\|u\|^2$ is real and nonzero, we conclude that the eigenvalue λ must be real too.

One consequence of this conclusion is that we can always find *real-valued* eigenvectors for symmetric matrices. Indeed, if we express u in terms of its real and imaginary vector components, say, as

$$u = p + jq, \quad p, q \in \mathbb{R}^N \tag{1.8}$$

Then, using (1.3) and the fact that λ is real, we conclude that it must hold:

$$Ap = \lambda p, \quad Aq = \lambda q \tag{1.9}$$

so that p and q are eigenvectors associated with λ.

Spectral theorem

A second important property of real symmetric matrices, one whose proof requires a more elaborate argument and is deferred to Appendix 1.A, is that such matrices always have a *full* set of orthonormal eigenvectors. That is, if $A \in \mathbb{R}^{N\times N}$ is symmetric, then there will exist a set of N orthonormal real eigenvectors $u_n \in \mathbb{R}^N$ satisfying

$$Au_n = \lambda_n u_n, \quad \|u_n\|^2 = 1, \quad u_n^\mathsf{T} u_m = 0 \;\text{ for } n \neq m \tag{1.10}$$

where all N eigenvalues $\{\lambda_n, n = 1, 2, \ldots, N\}$ are real, and all eigenvectors $\{u_n\}$ have unit norm and are orthogonal to each other. This result is known as the *spectral theorem*. For illustration purposes, assume A is 3×3. Then, the above statement asserts that there will exist three real orthonormal vectors $\{u_1, u_2, u_3\}$ and three real eigenvalues $\{\lambda_1, \lambda_2, \lambda_3\}$ such that

$$A \underbrace{\begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}}_{\triangleq\, U} = \underbrace{\begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix}}_{\triangleq\, \Lambda} \tag{1.11}$$

where we are introducing the matrices U and Λ for compactness of notation: U contains real eigenvectors for A and Λ is a diagonal matrix with the corresponding eigenvalues. Then, we can write (1.11) more compactly, as

$$AU = U\Lambda \tag{1.12}$$

However, the fact that the columns of U are orthogonal to each other and have unit norms implies that U satisfies the important normalization property:

$$UU^{\mathsf{T}} = I_N \quad \text{and} \quad U^{\mathsf{T}}U = I_N \tag{1.13}$$

That is, the product of U with U^{T} (or U^{T} with U) results in the identity matrix of size $N \times N$ — see Prob. 1.1. We say that U is an *orthogonal* matrix. Using this property and multiplying the matrix equality (1.12) by U^{T} from the right we get

$$A \underbrace{UU^{\mathsf{T}}}_{=I} = U\Lambda U^{\mathsf{T}} \tag{1.14}$$

We therefore conclude that every real symmetric matrix A can be expressed in the following spectral (or eigen-) decomposition form:

$$\boxed{A = U\Lambda U^{\mathsf{T}}} \qquad (\textbf{eigen-decomposition}) \tag{1.15a}$$

where, for general dimensions, the $N \times N$ matrices Λ and U are constructed from the eigenvalues and orthonormal eigenvectors of A as follows:

$$\Lambda = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_N\} \tag{1.15b}$$

$$U = \begin{bmatrix} u_1 & u_2 & \ldots & u_N \end{bmatrix} \tag{1.15c}$$

Rayleigh-Ritz ratio

There is a useful characterization of the smallest and largest eigenvalues of real symmetric matrices, known as the Rayleigh-Ritz ratio. Specifically, if $A \in \mathbb{R}^{N \times N}$ is symmetric, then it holds that for all vectors $x \in \mathbb{R}^N$:

$$\boxed{\lambda_{\min}\|x\|^2 \leq x^{\mathsf{T}}Ax \leq \lambda_{\max}\|x\|^2} \tag{1.16}$$

as well as

$$\lambda_{\min} = \min_{x\neq 0}\left(\frac{x^{\mathsf T}Ax}{x^{\mathsf T}x}\right) = \min_{\|x\|=1} x^{\mathsf T}Ax \tag{1.17a}$$

$$\lambda_{\max} = \max_{x\neq 0}\left(\frac{x^{\mathsf T}Ax}{x^{\mathsf T}x}\right) = \max_{\|x\|=1} x^{\mathsf T}Ax \tag{1.17b}$$

where $\{\lambda_{\min}, \lambda_{\max}\}$ denote the smallest and largest eigenvalues of A. The ratio $x^{\mathsf T}Ax/x^{\mathsf T}x$ is called the Rayleigh-Ritz ratio.

Proof of (1.16) and (1.17a)–(1.17b): Consider the eigen-decomposition (1.15a) and introduce the vector $y = U^{\mathsf T}x$ for any vector x. Then,

$$x^{\mathsf T}Ax = x^{\mathsf T}U\Lambda U^{\mathsf T}x = y^{\mathsf T}\Lambda y = \sum_{n=1}^{N}\lambda_n y_n^2 \tag{1.18}$$

with the $\{y_n\}$ denoting the individual entries of y. Now since the squared terms $\{y_n^2\}$ are nonnegative and the $\{\lambda_n\}$ are real, we get

$$\lambda_{\min}\left(\sum_{n=1}^{N} y_n^2\right) \le \sum_{n=1}^{N}\lambda_n y_n^2 \le \lambda_{\max}\left(\sum_{n=1}^{N} y_n^2\right) \tag{1.19}$$

or, equivalently,

$$\lambda_{\min}\|y\|^2 \le x^{\mathsf T}Ax \le \lambda_{\max}\|y\|^2 \tag{1.20}$$

Using the fact that U is orthogonal and, hence,

$$\|y\|^2 = y^{\mathsf T}y = x\underbrace{UU^{\mathsf T}}_{=I}x = \|x\|^2 \tag{1.21}$$

we conclude that (1.16) holds. The lower (upper) bound in (1.19) is achieved when x is chosen as the eigenvector $u_{\min}(u_{\max})$ corresponding to $\lambda_{\min}(\lambda_{\max})$.

■

Example 1.1 (**Quadratic curve**) Consider the two-dimensional function

$$g(r,s) = ar^2 + as^2 + 2brs, \quad r,s\in\mathbb{R} \tag{1.22}$$

We would like to determine the largest and smallest values that the function attains on the circle $r^2 + s^2 = 1$. One way to solve the problem is to recognize that $g(r,s)$ can be rewritten as

$$g(r,s) = \underbrace{\begin{bmatrix} r & s\end{bmatrix}}_{\triangleq\, x^{\mathsf T}}\underbrace{\begin{bmatrix} a & b\\ b & a\end{bmatrix}}_{\triangleq\, A}\underbrace{\begin{bmatrix} r\\ s\end{bmatrix}}_{=x} = x^{\mathsf T}Ax \tag{1.23}$$

We therefore want to determine the extreme values of the quadratic form $x^{\mathsf T}Ax$ under the constraint $\|x\| = 1$. According to (1.17a)–(1.17b), these values correspond to $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$. It can be easily verified that the eigenvalues of A are given by $\lambda(A) = \{a-b, a+b\}$ and, hence,

$$\lambda_{\min}(A) = \min\{a-b, a+b\}, \qquad \lambda_{\max}(A) = \max\{a-b, a+b\} \tag{1.24}$$

1.2 POSITIVE-DEFINITE MATRICES

An $N \times N$ real symmetric matrix A is said to be nonnegative-definite (also called positive semi-definite) if it satisfies the property:

$$v^{\mathsf{T}} A v \geq 0, \quad \text{for all column vectors } v \in \mathbb{R}^N \tag{1.25}$$

The matrix A is said to be positive-definite if $v^{\mathsf{T}} A v > 0$ for all $v \neq 0$. We denote a positive-definite matrix by writing $A > 0$ and a positive semi-definite matrix by writing $A \geq 0$.

Example 1.2 (**Diagonal matrices**) The notion of positive semi-definiteness is trivial for diagonal matrices. Consider the diagonal matrix

$$A = \text{diag}\{a_1, a_2, a_3\} \in \mathbb{R}^{3\times 3} \tag{1.26}$$

and let

$$v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \in \mathbb{R}^3 \tag{1.27}$$

denote an arbitrary vector. Then, some simple algebra shows that

$$v^{\mathsf{T}} A v \;=\; a_1 v_1^2 + a_2 v_2^2 + a_3 v_3^2 \tag{1.28}$$

This expression will be nonnegative for *any* v if, and only if, the entries a_n are all nonnegative. This is because if any a_n is negative, say a_2, then we can select a vector v with an entry v_2 that is large enough to result in a negative term $a_2 v_2^2$ that exceeds the contribution of the other two terms in the sum $v^{\mathsf{T}} A v$. Therefore, for a diagonal matrix to be positive semi-definite, it is necessary and sufficient that its diagonal entries be nonnegative. Likewise, a diagonal matrix A is positive definite if, and only if, its diagonal entries are positive. We cannot extrapolate and say that a general non-diagonal matrix A is positive semi-definite if all its entries are nonnegative; this conclusion is *not* true, as the next example shows.

Example 1.3 (**Non-diagonal matrices**) Consider the 2×2 matrix

$$A = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \tag{1.29}$$

This matrix is positive-definite. Indeed, pick any nonzero column vector $v \in \mathbb{R}^2$. Then,

$$\begin{aligned} v^{\mathsf{T}} A v &= \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \\ &= 3v_1^2 + 3v_2^2 - 2v_1 v_2 \\ &= (v_1 - v_2)^2 \;+\; 2v_1^2 \;+\; 2v_2^2 \\ &> 0, \quad \text{for any } v \neq 0 \end{aligned} \tag{1.30}$$

Among the several equivalent characterizations of positive-definite matrices, we note that an $N \times N$ real symmetric matrix A is positive-definite if, and only if, all its N eigenvalues are positive:

$$\boxed{A > 0 \iff \{\lambda_n > 0\}_{n=1}^{N}} \tag{1.31}$$

One proof relies on the use of the eigen-decomposition of A.

Proof of (1.31): We need to prove the statement in both directions. Assume initially that A is positive-definite and let us establish that all its eigenvalues are positive. Let $A = U\Lambda U^{\mathsf{T}}$ denote the spectral decomposition of A. Let also u_n denote the n-th column of U corresponding to the eigenvalue λ_n, i.e., $Au_n = \lambda_n u_n$ with $\|u_n\|^2 = 1$. If we multiply this equality from the left by u_n^{T} we get

$$u_n^{\mathsf{T}} A u_n = \lambda_n \|u_n\|^2 = \lambda_n > 0 \tag{1.32}$$

where the last inequality follows from the fact that $u^{\mathsf{T}} A u > 0$ for any nonzero vector u since A is assumed to be positive-definite. Therefore, $A > 0$ implies $\lambda_n > 0$ for $n = 1, 2, \ldots, N$.

Conversely, assume all $\lambda_n > 0$ and let us show that $A > 0$. Multiply the equality $A = U\Lambda U^{\mathsf{T}}$ by any *nonzero* vector v and its transpose, from right and left, to get

$$v^{\mathsf{T}} A v = v^{\mathsf{T}} U \Lambda U^{\mathsf{T}} v \tag{1.33}$$

Now introduce the real diagonal matrix

$$D \;\stackrel{\Delta}{=}\; \text{diag}\left\{\sqrt{\lambda_1},\; \sqrt{\lambda_2},\; \ldots,\; \sqrt{\lambda_n}\right\} \tag{1.34}$$

and the vector

$$s \;\stackrel{\Delta}{=}\; D U^{\mathsf{T}} v \tag{1.35}$$

The vector s is nonzero. This can be seen as follows. Let $w = U^{\mathsf{T}} v$. Then, the vectors v and w have the same Euclidean norm since

$$\|w\|^2 \;=\; w^{\mathsf{T}} w \;=\; v^{\mathsf{T}} \underbrace{U U^{\mathsf{T}}}_{=I} v \;=\; v^{\mathsf{T}} v \;=\; \|v\|^2 \tag{1.36}$$

It follows that the vector w is nonzero since v is nonzero. Now since $s = Dw$ and all entries of D are nonzero, we conclude that $s \neq 0$. Returning to (1.33), we get

$$v^{\mathsf{T}} A v = \|s\|^2 > 0 \tag{1.37}$$

for any nonzero v, which establishes that $A > 0$.

■

In a similar vein, we can show that

$$\boxed{A \geq 0 \iff \{\lambda_n \geq 0\}_{n=1}^{N}} \tag{1.38}$$

Example 1.4 **(Positive-definite matrix)** Consider again the 2×2 matrix from Example 1.3:

$$A = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \tag{1.39}$$

We established in that example from first principles that $A > 0$. Alternatively, we can determine the eigenvalues of A and verify that they are positive. The eigenvalues are the roots of the characteristic equation, $\det(\lambda I - A) = 0$, which leads to the quadratic equation $(\lambda - 3)^2 - 1 = 0$ so that $\lambda_1 = 4 > 0$ and $\lambda_2 = 2 > 0$.

A second useful property of positive-definite matrices is that they have positive determinants. To see this, recall first that for two square matrices A and B it holds that

$$\det(AB) = \det(A)\ \det(B) \tag{1.40}$$

That is, the determinant of the product is equal to the product of the determinants. Now starting with a positive-definite matrix A, and applying the above determinant formula to its eigen-decomposition (1.15a), we get

$$\det A = (\det U)\,(\det \Lambda)\,\left(\det U^{\mathsf{T}}\right) \tag{1.41}$$

But $UU^{\mathsf{T}} = I$ so that

$$(\det U)\,\left(\det U^{\mathsf{T}}\right) = 1 \tag{1.42}$$

and we conclude that

$$\det A = \det \Lambda = \prod_{n=1}^{N} \lambda_n \tag{1.43}$$

This result is actually general and holds for arbitrary square matrices A (the matrices do not need to be symmetric or positive-definite): the determinant of a matrix is always equal to the product of its eigenvalues (counting multiplicities) — see Prob. 1.2. Now, when the matrix A happens to be positive-definite, all its eigenvalues will be positive and, hence,

$$\boxed{A > 0 \implies \det A > 0} \tag{1.44}$$

Note that this statement goes in one direction only; the converse is not true.

1.3 RANGE SPACES AND NULLSPACES

Let A denote an $N \times M$ real matrix without any constraint on the relative sizes of N and M. When $N = M$, we say that A is a square matrix. Otherwise, when $N > M$, we say that A is a "tall" matrix and when $N < M$ we say that A is a "fat" matrix.

Definitions

The *column span* or the *range space* of A is defined as the set of all $N \times 1$ vectors q that can be generated by Ap, for all $M \times 1$ vectors p. We denote the column span of A by

$$\boxed{\mathcal{R}(A) \;\triangleq\; \left\{\text{set of all } q \in \mathbb{R}^N \text{ such that } q = Ap \text{ for some } p \in \mathbb{R}^M\right\}} \tag{1.45}$$

Likewise, the nullspace of A is the set of all $M \times 1$ vectors p that are annihilated by A, namely, that satisfy $Ap = 0$. We denote the nullspace of A by

$$\boxed{\mathcal{N}(A) \;\triangleq\; \left\{\text{set of all } p \in \mathbb{R}^M \text{ such that } Ap = 0\right\}} \tag{1.46}$$

The rank of a matrix A is defined as the number of linearly independent columns of A. It can be verified that, for any matrix A, the number of linearly independent columns is also equal to the number of linearly independent rows — see Prob. 1.5. It therefore holds that

$$\boxed{\text{rank}(A) \leq \min\{N, M\}} \tag{1.47}$$

That is, the rank of a matrix never exceeds its smallest dimension. A matrix is said to have *full rank* if

$$\text{rank}(A) = \min\{N, M\} \tag{1.48}$$

Otherwise, the matrix is said to be *rank deficient*.

If A is a square matrix (i.e., $N = M$), then rank deficiency is equivalent to a zero determinant, $\det A = 0$. Indeed, if A is rank deficient, then there exists a nonzero p such that $Ap = 0$. This means that $\lambda = 0$ is an eigenvalue of A so that its determinant must be zero.

Useful relations

One useful property that follows from the definition of range spaces and nullspaces is that any vector $z \in \mathbb{R}^N$ from the nullspace of $A^{\mathsf T}$ (not A) is orthogonal to any vector $q \in \mathbb{R}^N$ in the range space of A, i.e.,

$$\boxed{z \in \mathcal{N}(A^{\mathsf T}),\;\; q \in \mathcal{R}(A) \quad\Longrightarrow\quad z^{\mathsf T} q = 0} \tag{1.49}$$

Proof of (1.49): Indeed, $z \in \mathcal{N}(A^{\mathsf T})$ implies that $A^{\mathsf T} z = 0$ or, equivalently, $z^{\mathsf T} A = 0$. Now write $q = Ap$ for some p. Then, $z^{\mathsf T} q = z^{\mathsf T} Ap = 0$, as desired.

■

A second useful property is that the matrices $A^{\mathsf T}A$ and $A^{\mathsf T}$ have the *same* range space (i.e., they span the same space). Also, A and $A^{\mathsf T}A$ have the same nullspace, i.e.,

$$\boxed{\mathcal{R}(A^{\mathsf T}) = \mathcal{R}(A^{\mathsf T}A), \qquad \mathcal{N}(A) = \mathcal{N}(A^{\mathsf T}A)} \tag{1.50}$$

Proof of (1.50): Consider a vector $q \in \mathcal{R}(A^\mathsf{T}A)$, i.e., $q = A^\mathsf{T}Ap$ for some p. Define $r = Ap$, then $q = A^\mathsf{T}r$. This shows that $q \in \mathcal{R}(A^\mathsf{T})$ and we conclude that $\mathcal{R}(A^\mathsf{T}A) \subset \mathcal{R}(A^\mathsf{T})$. The proof of the converse statement requires more effort.

Consider a vector $q \in \mathcal{R}(A^\mathsf{T})$ and let us show by contradiction that $q \in \mathcal{R}(A^\mathsf{T}A)$. Assume, to the contrary, that q does not lie in $\mathcal{R}(A^\mathsf{T}A)$. This implies by (1.49) that there exists a vector z in the nullspace of $A^\mathsf{T}A$ that is not orthogonal to q, i.e., $A^\mathsf{T}Az = 0$ and $z^\mathsf{T}q \neq 0$. Now, if we multiply the equality $A^\mathsf{T}Az = 0$ by z^T from the left we obtain that $z^\mathsf{T}A^\mathsf{T}Az = 0$ or, equivalently, $\|Az\|^2 = 0$. Therefore, Az is necessarily the zero vector, $Az = 0$. But from $q \in \mathcal{R}(A^\mathsf{T})$ we have that $q = A^\mathsf{T}p$ for some p. Then, it must hold that $z^\mathsf{T}q = z^\mathsf{T}A^\mathsf{T}p = 0$, which contradicts $z^\mathsf{T}q \neq 0$. Therefore, we must have $q \in \mathcal{R}(A^\mathsf{T}A)$ and we conclude that $\mathcal{R}(A^\mathsf{T}) \subset \mathcal{R}(A^\mathsf{T}A)$.

The second assertion in (1.50) is more immediate. If $Ap = 0$ then $A^\mathsf{T}Ap = 0$ so that $\mathcal{N}(A) \subset \mathcal{N}(A^\mathsf{T}A)$. Conversely, if $A^\mathsf{T}Ap = 0$ then $p^\mathsf{T}A^\mathsf{T}Ap = \|Ap\|^2 = 0$ and we must have $Ap = 0$. That is, $\mathcal{N}(A^\mathsf{T}A) \subset \mathcal{N}(A)$. Combining both facts we conclude that $\mathcal{N}(A) = \mathcal{N}(A^\mathsf{T}A)$.

■

Normal equations

One immediate consequence of result (1.50) is that linear systems of equations of the following form:

$$\boxed{A^\mathsf{T}Ax = A^\mathsf{T}b} \qquad (\textbf{normal equations}) \tag{1.51}$$

are always consistent, i.e., they always have a solution x for *any* vector b. This is because $A^\mathsf{T}b$ belongs to $\mathcal{R}(A^\mathsf{T})$ and, therefore, also belongs to $\mathcal{R}(A^\mathsf{T}A)$. This type of linear systems of equations will appear as normal equations when we study least-squares problems later in Chapter 50 — see Eq. (50.25); the reason for the designation "normal equations" will be explained there. We can say more about the solution of such equations. For example, when the coefficient matrix $A^\mathsf{T}A$, which is always square regardless of the column and row dimensions of the $N \times M$ matrix A, happens to be invertible, then the normal equations (1.51) will have a unique solution given by

$$x = (A^\mathsf{T}A)^{-1}A^\mathsf{T}b \tag{1.52}$$

We explain further ahead in (1.58) that the matrix product $A^\mathsf{T}A$ will be invertible when the following two conditions hold: $N \geq M$ *and* A has full rank. In all other cases, the matrix product $A^\mathsf{T}A$ will be singular and will, therefore, have a nontrivial nullspace. Let p be any nonzero vector in the nullspace of $A^\mathsf{T}A$. We know from (1.50) that this vector also lies in the nullspace of A. Since we know that a solution x always exists for (1.51) then, by adding any such p to x, we obtain another solution. This is because:

$$\begin{aligned} A^\mathsf{T}A(x+p) &= A^\mathsf{T}Ax + A^\mathsf{T}Ap \\ &= A^\mathsf{T}Ax + 0 \\ &= A^\mathsf{T}Ax \\ &= A^\mathsf{T}b \end{aligned} \tag{1.53}$$

Knowing that there exist infinitely many vectors in $\mathcal{N}(A^\mathsf{T}A)$, e.g., any scaled multiple of p belongs to the same nullspace, we conclude that when $A^\mathsf{T}A$ is singular, there will exist infinitely many solutions to the normal equations (1.51). We therefore find that the normal equations (1.51) either have a unique solution (when $A^\mathsf{T}A$ is invertible) or infinitely many solutions (when $A^\mathsf{T}A$ is singular).

We can be more explicit about the latter case and verify that, when infinitely many solutions exist, they all differ by a vector in the nullspace of A. Indeed, assume $A^\mathsf{T}A$ is singular and let x_1 and x_2 denote two solutions to the normal equations (1.51). Then,

$$A^\mathsf{T}Ax_1 = A^\mathsf{T}b, \quad A^\mathsf{T}Ax_2 = A^\mathsf{T}b \tag{1.54}$$

Subtracting these two equalities we find that

$$A^\mathsf{T}A(x_1 - x_2) = 0 \tag{1.55}$$

which means that the difference $x_1 - x_2$ belongs to the nullspace of $A^\mathsf{T}A$ or, equivalently, to the nullspace of A in view of (1.50), namely,

$$x_1 - x_2 \in \mathcal{N}(A) \tag{1.56}$$

as claimed. We collect the results in the following statement for ease of reference.

Lemma 1.1. **(Solution of normal equations)** *Consider the normal system of equations $A^\mathsf{T}Ax = A^\mathsf{T}b$, where $A \in \mathbb{R}^{N\times M}$, $b \in \mathbb{R}^N$, and $x \in \mathbb{R}^M$. The following facts hold:*

(a) *A solution x always exists.*
(b) *The solution x is unique when $A^\mathsf{T}A$ is invertible (i.e., when $N \geq M$ and A has full rank). In this case, the solution is given by expression (1.52).*
(c) *There exist infinitely many solutions x when $A^\mathsf{T}A$ is singular.*
(d) *Under (c), any two solutions x_1 and x_2 will differ by a vector in $\mathcal{N}(A)$, i.e., (1.56) holds.*

The next result clarifies when the matrix product $A^\mathsf{T}A$ is invertible. Note in particular that the matrix $A^\mathsf{T}A$ is symmetric and nonnegative-definite; the latter property is because, for any nonzero x, it holds that

$$x^\mathsf{T}A^\mathsf{T}Ax \;=\; \|Ax\|^2 \;\geq 0 \tag{1.57}$$

Thus, let A be $N \times M$, with $N \geq M$ (i.e., A is a “tall” or square matrix). Then,

$$\boxed{A \text{ has full rank} \iff A^\mathsf{T}A \text{ is positive-definite}} \tag{1.58}$$

That is, every tall full rank matrix is such that the square matrix $A^\mathsf{T}A$ is invertible (actually, positive-definite).

Proof of (1.58): Assume first that A has full rank. This means that all columns of A are linearly independent, which in turn means that $Ax \neq 0$ for any nonzero x. Consequently, it holds that $\|Ax\|^2 > 0$, which is equivalent to $x^\mathsf{T}A^\mathsf{T}Ax > 0$ for any

$x \neq 0$. It follows that $A^\mathsf{T} A > 0$. Conversely, assume that $A^\mathsf{T} A > 0$. This means that $x^\mathsf{T} A^\mathsf{T} A x > 0$ for any nonzero x, which is equivalent to $\|Ax\|^2 > 0$ and, hence, $Ax \neq 0$. It follows that the columns of A are linearly independent so that A has full column rank.

■

In fact, when A has full rank, not only $A^\mathsf{T} A$ is positive-definite, but also any product of the form $A^\mathsf{T} B A$ for any symmetric positive-definite matrix B. Specifically, if $B > 0$, then

$$\boxed{A:\ N \times M,\ N \geq M,\ \text{ full rank } \quad \Longleftrightarrow \quad A^\mathsf{T} B A > 0} \tag{1.59}$$

Proof of (1.59): Assume first that A has full rank. This means that all columns of A are linearly independent, which in turn means that the vector $z = Ax \neq 0$ for any nonzero x. Now, since $B > 0$, it holds that $z^\mathsf{T} B z > 0$ and, hence, $x^\mathsf{T} A^\mathsf{T} B A x > 0$ for any nonzero x. It follows that $A^\mathsf{T} B A > 0$. Conversely, assume that $A^\mathsf{T} B A > 0$. This means that $x^\mathsf{T} A^\mathsf{T} B A x > 0$ for any nonzero x, which allows us to conclude, by contradiction, that A must be full rank. Indeed, assume not. Then, there should exist a nonzero vector p such that $Ap = 0$, which implies that $p^\mathsf{T} A^\mathsf{T} B A p = 0$. This conclusion contradicts the fact that $x^\mathsf{T} A^\mathsf{T} B A x > 0$ for any nonzero x. Therefore, A has full rank, as desired.

■

1.4 SCHUR COMPLEMENTS

There is a useful block triangularization formula that can often be used to facilitate the computation of matrix inverses or to reduce matrices to convenient block diagonal forms.

In this section we assume inverses exist whenever needed. Thus, consider a real block matrix:

$$S = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{1.60}$$

The Schur complement of A in S is denoted by Δ_A and is defined as the quantity:

$$\boxed{\Delta_A \triangleq D - CA^{-1}B} \tag{1.61}$$

Likewise, the Schur complement of D in S is denoted by Δ_D and is defined as

$$\boxed{\Delta_D \triangleq A - BD^{-1}C} \tag{1.62}$$

Block triangular factorizations

In terms of these Schur complements, it is easy to verify by direct calculation that the block matrix S can be factored in either of the following two useful forms:

$$\begin{aligned}\begin{bmatrix} A & B \\ C & D \end{bmatrix} &= \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & 0 \\ 0 & \Delta_A \end{bmatrix}\begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix} \\ &= \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} \Delta_D & 0 \\ 0 & D \end{bmatrix}\begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix}\end{aligned} \tag{1.63}$$

Two useful results that follow directly from these factorizations are the determinantal formulae:

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det A \ \det(D - CA^{-1}B) \tag{1.64a}$$

$$= \det D \ \det(A - BD^{-1}C) \tag{1.64b}$$

Block inversion formulas

Moreover, by inverting both sides of (1.63), we readily conclude that

$$\begin{aligned}\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} &= \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix}\begin{bmatrix} A^{-1} & 0 \\ 0 & \Delta_A^{-1} \end{bmatrix}\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \\ &= \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix}\begin{bmatrix} \Delta_D^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix}\begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix}\end{aligned} \tag{1.65}$$

where we used the fact that for block triangular matrices it holds

$$\begin{bmatrix} I & 0 \\ X & I \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -X & I \end{bmatrix}, \qquad \begin{bmatrix} I & X \\ 0 & I \end{bmatrix}^{-1} = \begin{bmatrix} I & -X \\ 0 & I \end{bmatrix} \tag{1.66}$$

If we expand expressions (1.65) we can also write

$$\begin{aligned}\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} &= \begin{bmatrix} A^{-1} + A^{-1}B\Delta_A^{-1}CA^{-1} & -A^{-1}B\Delta_A^{-1} \\ -\Delta_A^{-1}CA^{-1} & \Delta_A^{-1} \end{bmatrix} \\ &= \begin{bmatrix} \Delta_D^{-1} & -\Delta_D^{-1}BD^{-1} \\ -D^{-1}C\Delta_D^{-1} & D^{-1} + D^{-1}C\Delta_D^{-1}BD^{-1} \end{bmatrix}\end{aligned} \tag{1.67}$$

Matrix inertia and congruence

When the block matrix S is symmetric, its eigenvalues are real. We define the *inertia* of S as the triplet:

$$\text{In}\{S\} \ \stackrel{\Delta}{=} \ \Big\{I_+, I_-, I_0\Big\} \tag{1.68}$$

in terms of the integers:

$$I_+(S) = \text{the number of positive eigenvalues of } S \tag{1.69a}$$

$$I_-(S) = \text{the number of negative eigenvalues of } S \tag{1.69b}$$

$$I_0(S) = \text{the number of zero eigenvalues of } S \tag{1.69c}$$

Now, given a symmetric matrix S and any invertible matrix Q, the matrices S and QSQ^T are said to be *congruent*. An important result regarding congruent matrices is that congruence preserves inertia, i.e., it holds that

$$\boxed{\text{In}\{S\} = \text{In}\left\{QSQ^\mathsf{T}\right\}} \qquad (\textbf{congruence}) \tag{1.70}$$

so that the matrices S and QSQ^T will have the same number of positive, negative, and zero eigenvalues for any invertible Q. This result is known as Sylvester law of inertia.

Example 1.5 (**Inertia and Schur complements**) One immediate application of the above congruence result is the following characterization of the inertia of a matrix in terms of the inertia of its Schur complements. Thus, assume that S is symmetric with the block structure:

$$S = \begin{bmatrix} A & B \\ B^\mathsf{T} & D \end{bmatrix}, \quad \text{where } A = A^\mathsf{T} \text{ and } D = D^\mathsf{T} \tag{1.71}$$

Consider the corresponding block factorizations (1.63):

$$\begin{bmatrix} A & B \\ B^\mathsf{T} & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ B^\mathsf{T}A^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & \Delta_A \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix} \tag{1.72a}$$

$$= \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} \Delta_D & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ D^{-1}B^\mathsf{T} & I \end{bmatrix} \tag{1.72b}$$

in terms of the Schur complements:

$$\Delta_A = D - B^\mathsf{T}A^{-1}B, \qquad \Delta_D = A - BD^{-1}B^\mathsf{T} \tag{1.73}$$

The factorizations (1.72a)–(1.72b) have the form of congruence relations so that we must have

$$\text{In}\{S\} = \text{In}\left\{\begin{bmatrix} A & 0 \\ 0 & \Delta_A \end{bmatrix}\right\} \quad \text{and} \quad \text{In}\{S\} = \text{In}\left\{\begin{bmatrix} \Delta_D & 0 \\ 0 & D \end{bmatrix}\right\} \tag{1.74}$$

When S is positive-definite, all its eigenvalues are positive. Then, from the above inertia equalities, it follows that the matrices $\{A, \Delta_A, \Delta_D, D\}$ can only have positive eigenvalues. In other words, it must hold that

$$\boxed{S > 0 \iff A > 0 \text{ and } \Delta_A > 0} \tag{1.75a}$$

Likewise,

$$\boxed{S > 0 \iff D > 0 \text{ and } \Delta_D > 0} \tag{1.75b}$$

Example 1.6 (**A completion-of-squares formula**) Consider a quadratic expression of the form

$$J(x) = x^{\mathsf{T}}Ax - 2b^{\mathsf{T}}x + \alpha, \quad A \in \mathbb{R}^{M\times M}, \;\; b \in \mathbb{R}^{M\times 1}, \;\; \alpha \in \mathbb{R} \tag{1.76}$$

where A is assumed invertible and symmetric. We write $J(x)$ in the form

$$J(x) = \begin{bmatrix} x^{\mathsf{T}} & 1 \end{bmatrix} \begin{bmatrix} A & -b \\ -b^{\mathsf{T}} & \alpha \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix} \tag{1.77}$$

Usually, the block matrix

$$\begin{bmatrix} A & -b \\ -b^{\mathsf{T}} & \alpha \end{bmatrix} \tag{1.78}$$

is positive-definite in which case A is positive-definite and the Schur complement relative to it, namely, $\alpha - b^{\mathsf{T}}A^{-1}b$, is also positive. Next, we introduce the triangular factorization

$$\begin{bmatrix} A & -b \\ -b^{\mathsf{T}} & \alpha \end{bmatrix} = \begin{bmatrix} I_M & 0 \\ -b^{\mathsf{T}}A^{-1} & 1 \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & \alpha - b^{\mathsf{T}}A^{-1}b \end{bmatrix} \begin{bmatrix} I_M & -A^{-1}b \\ 0 & 1 \end{bmatrix} \tag{1.79}$$

and substitute it into (1.77) to get

$$\boxed{J(x) = (x - \widehat{x})^{\mathsf{T}}A(x - \widehat{x}) + (\alpha - b^{\mathsf{T}}A^{-1}b)} \tag{1.80}$$

where $\widehat{x} = A^{-1}b$. Decomposition (1.80) is referred to as a "sum-of-squares" expression since it is the sum of two positive terms.

Matrix inversion formula

For matrices of compatible dimensions, and invertible A, it holds that

$$\boxed{(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}} \tag{1.81}$$

This is a useful matrix identity that shows how the inverse of a matrix A is modified when it is perturbed by a product, BCD. The validity of the expression can be readily checked by multiplying both sides by $A + BCD$.

1.5 CHOLESKY FACTORIZATION

The Cholesky factorization of a positive-definite matrix is a useful computational tool and it can be motivated by means of the Schur decomposition results discussed above. Thus, consider an $M \times M$ symmetric positive-definite matrix A and partition it in the following manner

$$A = \begin{bmatrix} \alpha & b^{\mathsf{T}} \\ b & D \end{bmatrix} \tag{1.82}$$

where α is its leading diagonal entry, b is an $(M-1)\times 1$ column vector, and D has dimensions $(M-1)\times(M-1)$. The positive-definiteness of A guarantees

$\alpha > 0$ and $D > 0$. Using (1.72a), let us consider the following block factorization for A:

$$A = \begin{bmatrix} 1 & 0 \\ b/\alpha & I_{M-1} \end{bmatrix} \begin{bmatrix} \alpha & 0 \\ 0 & \Delta_\alpha \end{bmatrix} \begin{bmatrix} 1 & b^\mathsf{T}/\alpha \\ 0 & I_{M-1} \end{bmatrix} \tag{1.83a}$$

$$\Delta_\alpha = D - bb^\mathsf{T}/\alpha \tag{1.83b}$$

We can rewrite the factorization more compactly in the form:

$$A = \mathcal{L}_0 \begin{bmatrix} d(0) & \\ & \Delta_0 \end{bmatrix} \mathcal{L}_0^\mathsf{T} \tag{1.84}$$

where $\mathcal{L}_0$ is the lower-triangular matrix

$$\mathcal{L}_0 \triangleq \begin{bmatrix} 1 & 0 \\ b/\alpha & I_{M-1} \end{bmatrix} \tag{1.85}$$

and $d(0) = \alpha$, $\Delta_0 = \Delta_\alpha$. Observe that the first column of $\mathcal{L}_0$ is the first column of A normalized by the inverse of its leading diagonal entry. Moreover, the positive-definiteness of A guarantees $d(0) > 0$ and $\Delta_0 > 0$.

Expression (1.84) provides a factorization for A that consists of a lower-triangular matrix $\mathcal{L}_0$ followed by a *block-diagonal* matrix and an upper-triangular matrix. Now since Δ_0 is itself positive-definite, we can repeat the construction and introduce a similar factorization for it, which we denote by

$$\Delta_0 = L_1 \begin{bmatrix} d(1) & \\ & \Delta_1 \end{bmatrix} L_1^\mathsf{T} \tag{1.86}$$

for some lower-triangular matrix L_1 and where $d(1)$ is the leading diagonal entry of Δ_0. Moreover, Δ_1 is the Schur complement relative to $d(1)$ in Δ_0, and its dimensions are $(M-2) \times (M-2)$. In addition, the first column of L_1 coincides with the first column of Δ_0 normalized by the inverse of its leading diagonal entry. Also, the positive-definiteness of Δ_0 guarantees $d(1) > 0$ and $\Delta_1 > 0$. Substituting the above factorization for Δ_0 into the factorization for A we get

$$A = \mathcal{L}_0 \begin{bmatrix} 1 & \\ & L_1 \end{bmatrix} \begin{bmatrix} d(0) & & \\ & d(1) & \\ & & \Delta_1 \end{bmatrix} \begin{bmatrix} 1 & \\ & L_1^\mathsf{T} \end{bmatrix} \mathcal{L}_0^\mathsf{T} \tag{1.87}$$

But since the product of two lower-triangular matrices is also lower-triangular, we conclude that the product

$$\mathcal{L}_1 \triangleq \mathcal{L}_0 \begin{bmatrix} 1 & \\ & L_1 \end{bmatrix}$$

is lower-triangular and we denote it by $\mathcal{L}_1$. Using this notation, we write instead

$$A = \mathcal{L}_1 \begin{bmatrix} d(0) & & \\ & d(1) & \\ & & \Delta_1 \end{bmatrix} \mathcal{L}_1^\mathsf{T} \tag{1.88}$$

Clearly, the first column of $\mathcal{L}_1$ is the first column of $\mathcal{L}_0$ and the second column of $\mathcal{L}_1$ is formed from the first column of L_1.

We can proceed to factor Δ_1, which would lead to an expression of the form

$$A = \mathcal{L}_2 \begin{bmatrix} d(0) & & & \\ & d(1) & & \\ & & d(2) & \\ & & & \Delta_2 \end{bmatrix} \mathcal{L}_2^{\mathsf{T}} \tag{1.89}$$

where $d(2) > 0$ is the $(0,0)$ entry of Δ_1 and $\Delta_2 > 0$ is the Schur complement of $d(2)$ in Δ_1. Continuing in this fashion we arrive after $(M-1)$ Schur complementation steps at a factorization for A of the form

$$A = \mathcal{L}_{M-1} \mathcal{D} \mathcal{L}_{M-1}^{\mathsf{T}} \tag{1.90}$$

where $\mathcal{L}_{M-1}$ is $M \times M$ lower-triangular and $\mathcal{D}$ is $M \times M$ diagonal with positive entries $\{d(m)\}$. The columns of $\mathcal{L}_{M-1}$ are the successive leading columns of the Schur complements $\{\Delta_m\}$, normalized by the inverses of their leading diagonal entries. The diagonal entries of $\mathcal{D}$ coincide with these leading entries.

If we define $\bar{L} \stackrel{\Delta}{=} \mathcal{L}_{n-1}\mathcal{D}^{1/2}$, where $\mathcal{D}^{1/2}$ is a diagonal matrix with the positive square-roots of the $\{a(m)\}$, we obtain

$$\boxed{A = \bar{L}\bar{L}^{\mathsf{T}}} \quad \text{(lower-upper triangular factorization)} \tag{1.91}$$

In summary, this constructive argument shows that every positive-definite matrix can be factored as the product of a lower-triangular matrix with positive diagonal entries by its transpose. This factorization is known as the *Cholesky* factorization of A. Had we instead partitioned A as

$$A = \begin{bmatrix} B & b \\ b^{\mathsf{T}} & \beta \end{bmatrix} \tag{1.92}$$

where $\beta > 0$ is now a scalar, and had we used the block factorization (1.72b), we would have arrived at a similar factorization for A albeit one of the form:

$$\boxed{A = \bar{U}\bar{U}^{\mathsf{T}}} \quad \text{(upper-lower triangular factorization)} \tag{1.93}$$

where $\bar{U}$ is an upper-triangular matrix with positive diagonal entries. The two triangular factorizations are illustrated in Fig. 1.1.

Lemma 1.2. (Cholesky factorization) *Every positive-definite matrix A admits a unique factorization of either form $A = \bar{L}\bar{L}^{\mathsf{T}} = \bar{U}\bar{U}^{\mathsf{T}}$, where $\bar{L}$ ($\bar{U}$) is a lower (upper)-triangular matrix with positive entries along its diagonal.*

Proof: The existence of the factorizations was proved prior to the statement of the lemma. It remains to establish uniqueness. We show this for one of the factorizations. A similar argument applies to the other factorization. Thus, assume that

$$A = \bar{L}_1\bar{L}_1^{\mathsf{T}} = \bar{L}_2\bar{L}_2^{\mathsf{T}} \tag{1.94}$$

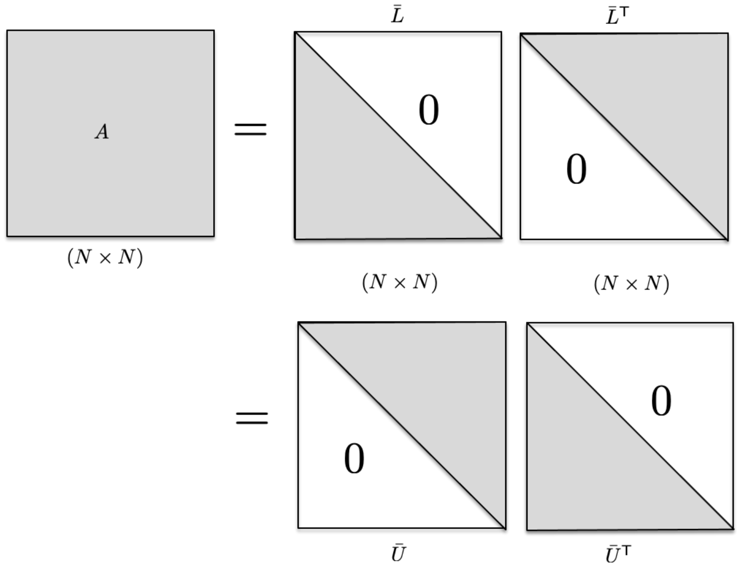

Figure 1.1 Two triangular factorizations for an $N \times N$ matrix A. The triangular factors $\{\bar{L}, \bar{U}\}$ are both $N \times N$ with positive entries on their main diagonal.

are two Cholesky factorizations for A. Then,

$$\bar{L}_2^{-1}\bar{L}_1 = \bar{L}_2^{\mathsf{T}}\bar{L}_1^{-\mathsf{T}} \tag{1.95}$$

where the compact notation $A^{-\mathsf{T}}$ stands for $[A^{\mathsf{T}}]^{-1}$. But since the inverse of a lower-triangular matrix is lower-triangular, and since the product of two lower-triangular matrices is also lower-triangular, we conclude that $\bar{L}_2^{-1}\bar{L}_1$ is lower-triangular. Likewise, the product $\bar{L}_2^{\mathsf{T}}\bar{L}_1^{-\mathsf{T}}$ is upper-triangular. Therefore, equality (1.95) will hold if, and only if, $\bar{L}_2^{-1}\bar{L}_1$ is diagonal, which means that

$$\bar{L}_1 = \bar{L}_2 D \tag{1.96}$$

for some diagonal matrix D. We want to show that D is the identity matrix. Indeed, it is easy to see from (1.94) that the $(0,0)$ entries of $\bar{L}_1$ and $\bar{L}_2$ must coincide so that the leading entry of D must be unity. This further implies from (1.96) that the first column of $\bar{L}_1$ should coincide with the first column of $\bar{L}_2$, so that using (1.94) again we conclude that the $(1,1)$ entries of $\bar{L}_1$ and $\bar{L}_2$ also coincide. Hence, the second entry of D is also unity. Proceeding in this fashion we conclude $D = I$.

■

Remark 1.1. **(Triangular factorization)** We also conclude from the discussion in this section that every positive-definite matrix A admits a unique factorization of either form $A = LDL^{\mathsf{T}} = UD_uU^{\mathsf{T}}$, where L (U) is a lower (upper)-triangular matrix with *unit* diagonal entries, and D and D_u are diagonal matrices with positive entries.

■

1.6 QR DECOMPOSITION

The QR decomposition of a matrix is a very useful tool; for example, it can be used to derive numerically robust implementations for the solution of normal equations of the form (1.51) — see the comments at the end of the chapter following (1.216) and also future Prob. 50.5. It can also be used to replace a collection of vectors by an orthonormal basis.

Consider an $N \times M$ real matrix A with $N \geq M$. We denote the individual columns of A by $\{h_m,\ m = 1, 2, \ldots, M\}$:

$$A = \begin{bmatrix} h_1 & h_2 & \ldots & h_M \end{bmatrix}, \quad h_m \in \mathbb{R}^N \tag{1.97}$$

The column span, $\mathcal{R}(A)$, is the result of all linear combinations of these columns. The vectors $\{h_m\}$ are not generally orthogonal to each other. They can, however, be converted into an orthonormal set of vectors, which we denote by $\{q_m,\ m = 1, 2, \ldots, M\}$ and which will span the same $\mathcal{R}(A)$. This objective can be achieved by means of the *Gram-Schmidt procedure*. It is an iterative procedure that starts by setting:

$$q_1 = h_1/\|h_1\|, \quad r_1 \stackrel{\Delta}{=} h_1 \tag{1.98}$$

and then repeats for $m = 2, \ldots, M$:

$$r_m = h_m - \sum_{j=1}^{m-1} \left(h_m^\mathsf{T} q_j\right) q_j \tag{1.99a}$$

$$q_m = r_m/\|r_m\| \tag{1.99b}$$

By iterating this construction, we end up expressing each column h_m as a linear combination of the vectors $\{q_1, q_2, \ldots, q_m\}$ as follows:

$$h_m = (h_m^\mathsf{T} q_1) q_1 + (h_m^\mathsf{T} q_2) q_2 + \ldots + (h_m^\mathsf{T} q_{m-1}) q_{m-1} + \|r_m\|\, q_m \tag{1.100}$$

If we collect the coefficients of this linear combination, for all $m = 1, 2, \ldots, M$, into the columns of an $M \times M$ upper-triangular matrix R:

$$R \stackrel{\Delta}{=} \begin{bmatrix} \|r_1\| & h_2^\mathsf{T} q_1 & h_3^\mathsf{T} q_1 & \ldots & h_M^\mathsf{T} q_1 \\ & \|r_2\| & h_3^\mathsf{T} q_2 & \ldots & h_M^\mathsf{T} q_2 \\ & & \|r_3\| & & \vdots \\ & & & \ddots & h_M^\mathsf{T} q_{M-1} \\ & & & & \|r_M\| \end{bmatrix} \tag{1.101}$$

we conclude from (1.100) that

$$A = \widehat{Q} R \tag{1.102}$$

where $\widehat{Q}$ is the $N \times M$ matrix with orthonormal columns $\{q_m\}$, i.e.,

$$\widehat{Q} = \begin{bmatrix} q_1 & q_2 & \ldots & q_M \end{bmatrix} \tag{1.103}$$

with

$$q_m^{\mathsf{T}} q_s = \begin{cases} 0, & m \neq s \\ 1, & m = s \end{cases} \tag{1.104}$$

When A has full rank, i.e., when A has rank M, all the diagonal entries of R will be positive. The factorization $A = \widehat{Q}R$ is referred to as the *reduced* QR decomposition of A, and it simply amounts to the orthonormalization of the columns of A.

It is often more convenient to employ the *full* QR decomposition of $A \in \mathbb{R}^{N\times M}$, as opposed to its reduced decomposition. The full decomposition is obtained by appending $N - M$ orthonormal columns to $\widehat{Q}$ so that it becomes an orthogonal $N \times N$ (square) matrix Q. We also append rows of zeros to R so that (1.102) becomes — see Fig. 1.2:

$$\boxed{A = Q \begin{bmatrix} R \\ 0 \end{bmatrix}, \quad Q\,(N\times N),\; R\,(M\times M)} \tag{1.105}$$

where

$$Q = \begin{bmatrix} \widehat{Q} & q_{M+1} & \ldots & q_N \end{bmatrix}, \quad Q^{\mathsf{T}} Q = I_N \tag{1.106}$$

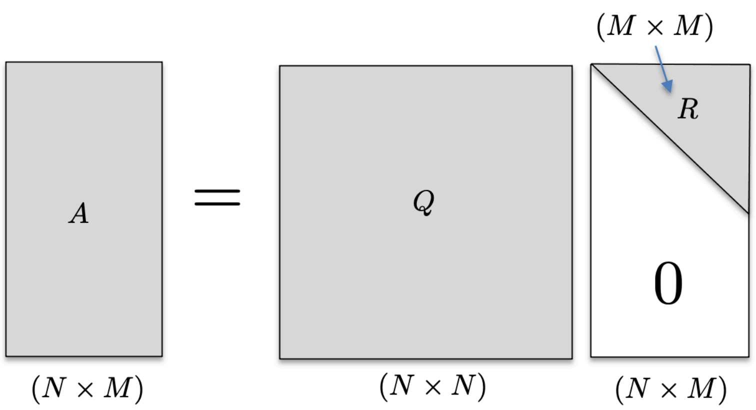

Figure 1.2 Full QR decomposition of an $N \times M$ matrix A, where Q is $N \times N$ orthogonal and R is $M \times M$ upper triangular.

Example 1.7 **(Cholesky and QR factorizations)** Consider a full rank $N \times M$ matrix A with $N \geq M$ and introduce its QR decomposition (1.105). Then, $A^{\mathsf{T}}A$ is positive-definite and its Cholesky factorization is given by

$$A^{\mathsf{T}} A = R^{\mathsf{T}} R \tag{1.107}$$

1.7 SINGULAR VALUE DECOMPOSITION

The singular value decomposition (SVD) of a matrix is another powerful tool that is useful for both analytical and numerical purposes. It enables us to represent any matrix (square or not, invertible or not, symmetric or not) as the product of three matrices with special and desirable properties: two of the matrices are orthogonal and the third matrix is composed of a diagonal matrix and a zero block.

Definition

The SVD of a real matrix A states that if A is $N \times M$, then there exist an $N \times N$ orthogonal matrix U ($UU^{\mathsf{T}} = I_N$), an $M \times M$ orthogonal matrix V ($VV^{\mathsf{T}} = I_M$), and a diagonal matrix Σ with nonnegative entries such that:

(a) If $N \leq M$, then Σ is $N \times N$ and

$$\boxed{A = U \begin{bmatrix} \Sigma & 0 \end{bmatrix} V^{\mathsf{T}}, \quad A \in \mathbb{R}^{N \times M}, \quad N \leq M} \tag{1.108a}$$

(b) If $N \geq M$, then Σ is $M \times M$ and

$$\boxed{A = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^{\mathsf{T}}, \quad A \in \mathbb{R}^{N \times M}, \quad N \geq M} \tag{1.108b}$$

Observe that U and V are square matrices, while the central matrix in (1.108a)–(1.108b) has the dimensions of A. Both factorizations are illustrated in Fig. 1.3. The diagonal entries of Σ are called the *singular values* of A and are ordered in decreasing order, say,

$$\Sigma = \text{diag}\Big\{\sigma_1, \sigma_2, \ldots, \sigma_r, 0, \ldots, 0\Big\} \tag{1.109}$$

with

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0 \tag{1.110}$$

If Σ has r nonzero diagonal entries then A has rank r. The columns of U and V are called the left and right *singular vectors* of A, respectively. The ratio of the largest to smallest singular value of A is called the *condition number* of A and is denoted by

$$\kappa(A) \overset{\Delta}{=} \sigma_1/\sigma_r \tag{1.111}$$

One constructive proof for the SVD is given in Appendix 1.B.

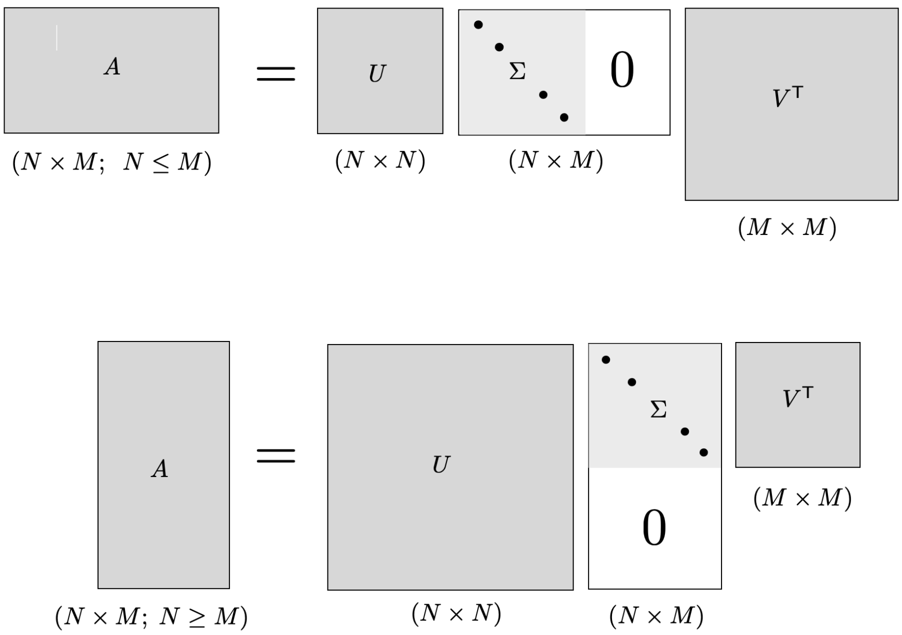

Figure 1.3 Singular value decompositions of an $N \times M$ matrix A for both cases when $N \geq M$ and $N \leq M$.

Pseudo Inverses

The pseudo-inverse of a matrix is a generalization of the concept of inverses for square invertible matrices; it is defined for matrices that need not be invertible or even square.

Given an $N \times M$ matrix A of rank r, its pseudo-inverse is defined as the unique $M \times N$ matrix $A^\dagger$ that satisfies the following four requirements:

$$\textbf{(i)}\ \ AA^\dagger A = A \tag{1.112a}$$

$$\textbf{(ii)}\ \ A^\dagger A A^\dagger = A^\dagger \tag{1.112b}$$

$$\textbf{(iii)}\ \ (AA^\dagger)^\mathsf{T} = AA^\dagger \tag{1.112c}$$

$$\textbf{(iv)}\ \ (A^\dagger A)^\mathsf{T} = A^\dagger A \tag{1.112d}$$

The SVD of A can be used to determine its pseudo-inverse as follows. Introduce the matrix

$$\Sigma^\dagger = \text{diagonal}\left\{\sigma_1^{-1}, \sigma_2^{-1}, \ldots, \sigma_r^{-1}, 0, \ldots, 0\right\} \tag{1.113}$$

That is, we invert the nonzero entries of Σ and keep the zero entries unchanged.

(a) When $N \leq M$, we define

$$A^\dagger = V \begin{bmatrix} \Sigma^\dagger \\ 0 \end{bmatrix} U^\mathsf{T} \tag{1.114}$$

(b) When $N \geq M$, we define

$$A^{\dagger} = V \left[\begin{array}{cc} \Sigma^{\dagger} & 0 \end{array} \right] U^{\mathsf{T}} \tag{1.115}$$

It can be verified that these expressions for $A^{\dagger}$ satisfy the four defining properties (1.112a)–(1.112d) listed above.

Example 1.8 **(Full rank matrices)** It can also be verified, by replacing A by its SVD in the expressions below, that when $A \in \mathbb{R}^{N\times M}$ has *full rank*, its pseudo-inverse is given by the following expressions:

$$A^{\dagger} = A^{\mathsf{T}}(AA^{\mathsf{T}})^{-1}, \quad \text{when } N \leq M \tag{1.116a}$$
$$A^{\dagger} = (A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}, \quad \text{when } N \geq M \tag{1.116b}$$

1.8 SQUARE-ROOT MATRICES

One useful concept in matrix analysis is that of the *square-root matrix*. Although square-roots can be defined for nonnegative-definite matrices, it is sufficient for our purposes to focus on positive-definite matrices. Thus, consider an $N \times N$ positive-definite matrix A and introduce its eigen-decomposition

$$A = U\Lambda U^{\mathsf{T}} \tag{1.117}$$

where Λ is an $N \times N$ diagonal with positive entries and U is $N \times N$ orthogonal:

$$UU^{\mathsf{T}} = U^{\mathsf{T}}U = I_N \tag{1.118}$$

Let $\Lambda^{1/2}$ denote the diagonal matrix whose entries are the positive square-roots of the diagonal entries of Λ. Then, we can rewrite (1.117) as

$$A = \left(U\Lambda^{1/2}\right)\left(U\Lambda^{1/2}\right)^{\mathsf{T}} \tag{1.119}$$

which expresses A as the product of an $N \times N$ matrix and its transpose, namely,

$$A = XX^{\mathsf{T}}, \text{ with } X = U\Lambda^{1/2} \tag{1.120}$$

We say that X is *a* square-root for A.

Definition 1.1. (Square-root factors) *A square-root of an $N \times N$ positive-definite matrix A is any $N \times N$ matrix X satisfying $A = XX^{\mathsf{T}}$. The square-root is said to be symmetric if $X = X^{\mathsf{T}}$ in which case $A = X^2$.*

The construction prior to the definition exhibits one possible choice for X, namely, $X = U\Lambda^{1/2}$, in terms of the eigenvectors and eigenvalues of A. However, square-root factors are not unique. If we consider the above X and multiply it by any orthogonal matrix Θ, say, $\bar{X} = X\Theta$ where $\Theta\Theta^{\mathsf{T}} = I_N$, then $\bar{X}$ is also a square-root factor for A since

$$\bar{X}\bar{X}^{\mathsf{T}} \;=\; X\underbrace{\Theta\Theta^{\mathsf{T}}}_{=I}X^{\mathsf{T}} = XX^{\mathsf{T}} = A \tag{1.121}$$

In particular, for the same matrix $A = U\Lambda U^{\mathsf{T}}$, the matrix $X = U\Lambda^{1/2}U^{\mathsf{T}}$ is also a square-root for A. And, this particular square-root factor happens to be symmetric. This argument shows that every symmetric positive-definite matrix A admits a *symmetric* square-root factor for which we can write $A = X^2$.

Notation

It is customary to use the notation $A^{1/2}$ to refer to a square-root of a matrix A and, therefore, we write

$$A = A^{1/2}\left(A^{1/2}\right)^{\mathsf{T}} \tag{1.122a}$$

It is also customary to employ the compact notation

$$A^{\mathsf{T}/2} \;\triangleq\; \left(A^{1/2}\right)^{\mathsf{T}}, \quad A^{-1/2} \;\triangleq\; \left(A^{1/2}\right)^{-1}, \quad A^{-\mathsf{T}/2} \;\triangleq\; \left(A^{1/2}\right)^{-\mathsf{T}} \tag{1.122b}$$

so that

$$\boxed{A = A^{1/2}A^{\mathsf{T}/2}, \qquad A^{-1} = A^{-\mathsf{T}/2}A^{-1/2}} \tag{1.122c}$$

Cholesky factor

One of the most widely used square-root factors of a positive-definite matrix is its Cholesky factor. Recall that we showed in Sec. 1.5 that every positive-definite matrix A admits a *unique* triangular factorization of the form $A = \bar{L}\bar{L}^{\mathsf{T}}$, where $\bar{L}$ is a lower-triangular matrix with positive entries on its diagonal. We could also consider the alternative factorization $A = \bar{U}\bar{U}^{\mathsf{T}}$ in terms of an upper triangular matrix $\bar{U}$. Comparing these forms with the defining relation $A = XX^{\mathsf{T}}$, we conclude that $\bar{L}$ and $\bar{U}$ are valid choices for square-root factors of A. When one refers to the square-root of a matrix, it is generally meant its (lower or upper triangular) Cholesky factor. This choice has two advantages in relation to other square-root factors: it is triangular and is uniquely defined (i.e., there is no other triangular square-root factor with positive diagonal entries).

Example 1.9 (**Basis rotation**) The following is an important matrix result that is critical to the development of algorithms that compute and propagate square-root factors. Consider two $N \times M$ ($N \leq M$) matrices A and B. Then $AA^{\mathsf{T}} = BB^{\mathsf{T}}$ if, and only if, there exists an $M \times M$ orthogonal matrix Θ such that $A = B\Theta$.

Proof: One direction is obvious. If $A = B\Theta$, for some orthogonal matrix Θ, then

$$AA^{\mathsf{T}} = (B\Theta)(B\Theta)^{\mathsf{T}} = B(\Theta\Theta^{\mathsf{T}})B^{\mathsf{T}} = BB^{\mathsf{T}} \tag{1.123}$$

One proof for the converse implication follows by using the singular value decompositions of A and B:

$$A = U_A \begin{bmatrix} \Sigma_A & 0 \end{bmatrix} V_A^{\mathsf{T}}, \quad B = U_B \begin{bmatrix} \Sigma_B & 0 \end{bmatrix} V_B^{\mathsf{T}} \tag{1.124}$$

where U_A and U_B are $N \times N$ orthogonal matrices, V_A and V_B are $M \times M$ orthogonal matrices, and Σ_A and Σ_B are $N \times N$ diagonal matrices with nonnegative entries. The squares of the diagonal entries of Σ_A (Σ_B) are the eigenvalues of AA^{T} (BB^{T}). Moreover, U_A (U_B) are constructed from an orthonormal basis for the right eigenvectors of AA^{T} (BB^{T}). Hence, it follows from the identity $AA^{\mathsf{T}} = BB^{\mathsf{T}}$ that $\Sigma_A = \Sigma_B$ and $U_A = U_B$. Let $\Theta = V_B V_A^{\mathsf{T}}$. Then, it holds that $\Theta\Theta^{\mathsf{T}} = I$ and $B\Theta = A$.

■

1.9 KRONECKER PRODUCTS

Let $A = [a_{ij}]_{i,j=1}^{N}$ and $B = [b_{ij}]_{i,j=1}^{M}$ be $N \times N$ and $M \times M$ real-valued matrices, respectively, whose individual $(i,j)-$th entries are denoted by a_{ij} and b_{ij}. Their Kronecker product is denoted by $\mathcal{K} = A \otimes B$ and is defined as the $NM \times NM$ matrix whose entries are given by:

$$\mathcal{K} \triangleq A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1N}B \\ a_{21}B & a_{22}B & \dots & a_{2N}B \\ \vdots & & \vdots & \\ a_{N1}B & a_{N2}B & \dots & a_{NN}B \end{bmatrix} \tag{1.125}$$

In other words, each scalar entry a_{ij} of A is replaced by a block quantity that is equal to a scaled multiple of B, namely, $a_{ij}B$.

1.9.1 Properties

Let $\{\lambda_i(A), i = 1, \dots, N\}$ and $\{\lambda_j(B), j = 1, \dots, M\}$ denote the eigenvalues of A and B, respectively. Then, the eigenvalues of $A \otimes B$ will consist of all nm product combinations $\{\lambda_i(A)\lambda_j(B)\}$. A similar conclusion holds for the singular values of $A \otimes B$ in relation to the singular values of the individual matrices A and B, which we denote by $\{\sigma_i(A), \sigma_j(B)\}$. Table 1.1 lists several well-known properties of Kronecker products for matrices $\{A, B, C, D\}$ of compatible dimensions and column vectors $\{x, y\}$. The last three properties involve the trace and vec operations: the trace of a square matrix is the sum of its diagonal elements, while the vec operation transforms a matrix into a vector by stacking the columns of the matrix on top of each other.

Example 1.10 **(Derivation of select properties)** Property (2) in Table 1.1 follows by direct calculation from the definition of Kronecker products. Property (4) follows by using property (2) to note that

$$(A \otimes B)(A^{-1} \otimes B^{-1}) = I_N \otimes I_M = I_{NM} \tag{1.126}$$

Table 1.1 Properties for the Kronecker product (1.125).

	relation	property
1.	$(A+B)\otimes C = (A\otimes C)\ +\ (B\otimes C)$	distributive property
2.	$(A\otimes B)(C\otimes D) = (AC\otimes BD)$	multiplication property
3.	$(A\otimes B)^{\mathsf{T}} = A^{\mathsf{T}}\otimes B^{\mathsf{T}}$	transposition property
4.	$(A\otimes B)^{-1} = A^{-1}\otimes B^{-1}$	inversion property
5.	$(A\otimes B)^{\ell} = A^{\ell}\otimes B^{\ell}$, integer ℓ	exponentiation property
6.	$\{\lambda(A\otimes B)\} = \{\lambda_i(A)\lambda_j(B)\}_{i=1,j=1}^{N,M}$	eigenvalues
7.	$\{\sigma(A\otimes B)\} = \{\sigma_i(A)\sigma_j(B)\}_{i=1,j=1}^{N,M}$	singular values
8.	$\det(A\otimes B) = (\det A)^M(\det B)^N$	determinant property
9.	$\mathrm{Tr}(A\otimes B) = \mathrm{Tr}(A)\mathrm{Tr}(B)$	trace of Kronecher product
10.	$\mathrm{Tr}(AB) = \Big(\mathrm{vec}(B^{\mathsf{T}})\Big)^{\mathsf{T}}\mathrm{vec}(A)$	trace of matrix product
11.	$\mathrm{vec}(ACB) = (B^{\mathsf{T}}\otimes A)\mathrm{vec}(C)$	vectorization property
12.	$\mathrm{vec}(xy^{\mathsf{T}}) = y\otimes x$	vectorization of outer product

Property (6) follows from property (2) by choosing C as a right eigenvector for A and D as a right eigenvector for B, say, $C = q_i$ and $D = p_j$ where

$$Aq_i = \lambda_i(A)q_i, \qquad Bp_j = \lambda_j(B)p_j \tag{1.127}$$

Then,

$$(A\otimes B)(q_i\otimes p_j) = \lambda_i(A)\lambda_j(B)(q_i\otimes p_j) \tag{1.128}$$

which shows that $(q_i\otimes p_j)$ is an eigenvector of $(A\otimes B)$ with eigenvalue $\lambda_i(A)\lambda_j(B)$. Property (9) follows from property (6) for square matrices since

$$\mathrm{Tr}(A) = \sum_{i=1}^{N}\lambda_i(A), \qquad \mathrm{Tr}(B) = \sum_{j=1}^{M}\lambda_j(B) \tag{1.129}$$

and, therefore,

$$\begin{aligned}\mathrm{Tr}(A)\mathrm{Tr}(B) &= \left(\sum_{i=1}^{N}\lambda_i(A)\right)\left(\sum_{j=1}^{M}\lambda_j(B)\right)\\ &= \sum_{i=1}^{N}\sum_{j=1}^{M}\lambda_i(A)\lambda_j(B)\\ &= \mathrm{Tr}(A\otimes B)\end{aligned} \tag{1.130}$$

Property (8) also follows from property (6) since

$$\begin{aligned}\det(A\otimes B) &= \prod_{i=1}^{N}\prod_{j=1}^{M}\lambda_i(A)\lambda_j(B)\\ &= \left(\prod_{i=1}^{N}\lambda_i(A)\right)^M\left(\prod_{j=1}^{M}\lambda_j(B)\right)^N\\ &= (\det A)^M(\det B)^N\end{aligned} \tag{1.131}$$

Property (11) follows from the definition of Kronecker products and from noting that, for any two column vectors x and y, the vec representation of the rank one matrix xy^T is $y \otimes x$, i.e., $\text{vec}(xy^\mathsf{T}) = y \otimes x$, which is property (12). Finally, property (3) follows from the definition of Kronecker products.

Example 1.11 (**Discrete-time Lyapunov equations**) Consider $N \times N$ matrices X, A, and Q, where Q is symmetric and non-negative definite. The matrix X is said to satisfy a discrete-time Lyapunov equation, also called a Stein equation, if

$$\boxed{X - A^\mathsf{T} X A = Q} \tag{1.132}$$

Let $\lambda_k(A)$ denote any of the eigenvalues of A. We say that A is a *stable* matrix when all of its eigenvalues lie strictly inside the unit disc (i.e., their magnitudes are strictly less than one). Using properties of the Kronecker product, it can be verified that the following important facts hold:

(a) The solution X of (1.132) is unique if, and only if, $\lambda_k(A)\lambda_\ell(A) \neq 1$ for all $k, \ell = 1, 2, \ldots, N$. In this case, the unique solution X is symmetric.

(b) When A is stable, the solution X is unique, symmetric, and nonnegative-definite. Moreover, it admits the series representation:

$$X = \sum_{n=0}^{\infty} (A^\mathsf{T})^n Q A^n \tag{1.133}$$

Proof: We call upon property (11) from Table 1.1 and apply the vec operation to both sides of (1.132) to get

$$(I - A^\mathsf{T} \otimes A^\mathsf{T})\text{vec}(X) = \text{vec}(Q) \tag{1.134}$$

This linear system of equations has a unique solution, $\text{vec}(X)$, if, and only if, the coefficient matrix, $I - A^\mathsf{T} \otimes A^\mathsf{T}$, is nonsingular. This latter condition requires $\lambda_k(A)\lambda_\ell(A) \neq 1$ for all $k, \ell = 1, 2, \ldots, N$. When A is stable, all of its eigenvalues lie strictly inside the unit disc and this uniqueness condition is automatically satisfied. If we transpose both sides of (1.132) we find that X^T satisfies the same Lyapunov equation as X and, hence, by uniqueness, we must have $X = X^\mathsf{T}$. Finally, let $F = A^\mathsf{T} \otimes A^\mathsf{T}$. When A is stable, the matrix F is also stable by property (6) from Table 1.1. In this case, the matrix inverse $(I - F)^{-1}$ admits the series expansion

$$(I - F)^{-1} = I + F + F^2 + F^3 + \ldots \tag{1.135}$$

so that using (1.134) we have

$$\begin{aligned}
\text{vec}(X) &= (I - F)^{-1}\text{vec}(Q) \\
&= \sum_{n=0}^{\infty} F^n \,\text{vec}(Q) \\
&= \sum_{n=0}^{\infty} \left((A^\mathsf{T})^n \otimes (A^\mathsf{T})^n\right) \text{vec}(Q) \\
&= \sum_{n=0}^{\infty} \text{vec}\left((A^\mathsf{T})^n Q A^n\right)
\end{aligned} \tag{1.136}$$

from which we deduce the series representation (1.133). The last equality in (1.136) follows from property (11) in Table 1.1.

Example 1.12 **(Continuous-time Lyapunov equations)** We extend the analysis of Example 1.11 to the following continuous-time Lyapunov equation (also called a Sylvester equation):

$$\boxed{XA^{\mathsf{T}} + AX + Q = 0} \tag{1.137}$$

where Q continues to be symmetric and nonnegative definite. In the continuous-time case, a stable matrix A is one whose eigenvalues lie in the open left-half plane (i.e., they have strictly negative real parts). The following facts hold:

(a) The solution X of (1.137) is unique if, and only if, $\lambda_k(A) + \lambda_\ell(A) \neq 0$ for all $k, \ell = 1, 2, \ldots, N$. In this case, the unique solution X is symmetric.

(b) When A is stable (i.e., all its eigenvalues lie in the open left-half plane), the solution X is unique, symmetric, and nonnegative-definite. Moreover, it admits the integral representation

$$X = \int_0^\infty e^{At} Q e^{A^{\mathsf{T}} t} dt \tag{1.138}$$

where the notation e^{At} refers to the matrix exponential function evaluated at At. By definition, this function is equal to the following series representation:

$$e^{At} = I_N + \frac{1}{1!}At + \frac{1}{2!}A^2t^2 + \frac{1}{3!}A^3t^3 + \ldots \tag{1.139}$$

Proof: We use property (11) from Table 1.1 and apply the vec operation to both sides of (1.137) to get

$$\Big[(A^{\mathsf{T}} \otimes I) + (I \otimes A)\Big] \operatorname{vec}(X) = -\operatorname{vec}(Q) \tag{1.140}$$

This linear system of equations has a unique solution, $\operatorname{vec}(X)$, if, and only if, the coefficient matrix, $(A^{\mathsf{T}} \otimes I) + (I \otimes A)$, is nonsingular. This latter condition requires $\lambda_k(A) + \lambda_\ell(A) \neq 0$ for all $k, \ell = 1, 2, \ldots, N$. To see this, let $F = (A^{\mathsf{T}} \otimes I) + (I \otimes A)$ and let us verify that the eigenvalues of F are given by all linear combinations $\lambda_k(A) + \lambda_\ell(A)$. Consider the eigenvalue-eigenvector pairs $Ax_k = \lambda_k(A)x_k$ and $A^{\mathsf{T}}y_\ell = \lambda_\ell(A)y_\ell$. Then, using property (2) from Table 1.1 for Kronecker products we get

$$\begin{aligned}
F(y_\ell \otimes x_k) &= \Big[(A^{\mathsf{T}} \otimes I) + (I \otimes A)\Big](y_\ell \otimes x_k) \\
&= (A^{\mathsf{T}}y_\ell \otimes x_k) + (y_\ell \otimes Ax_k) \\
&= \lambda_\ell(A)(y_\ell \otimes x_k) + \lambda_k(A)(y_\ell \otimes x_k) \\
&= (\lambda_k(A) + \lambda_\ell(A))(y_\ell \otimes x_k)
\end{aligned} \tag{1.141}$$

so that the vector $(y_\ell \otimes x_k)$ is an eigenvector for F with eigenvalue $\lambda_k(A) + \lambda_\ell(A)$, as claimed. If we now transpose both sides of (1.137) we find that X^{T} satisfies the same Lyapunov equation as X and, hence, by uniqueness, we must have $X = X^{\mathsf{T}}$. Moreover, it follows from the integral representation (1.138) that X is nonnegative-definite since $Q \geq 0$ and $e^{A^{\mathsf{T}}t} = (e^{At})^{\mathsf{T}}$. To establish the integral representation we verify that it satisfies the Sylvester equation (1.137) so that, by uniqueness, the solution X should agree with it. Thus, let

$$Y \triangleq \int_0^\infty e^{At} Q e^{A^{\mathsf{T}} t} dt \tag{1.142}$$

and note that — refer to Prob. 1.20:

$$\begin{aligned} AY + YA^{\mathsf{T}} &= \int_0^\infty \left(Ae^{At}Qe^{A^{\mathsf{T}}t} + e^{At}Qe^{A^{\mathsf{T}}t}A^{\mathsf{T}}\right)dt \\ &= \int_0^\infty \frac{d}{dt}\left(e^{At}Qe^{A^{\mathsf{T}}t}\right)dt \\ &= e^{At}Qe^{A^{\mathsf{T}}t}\Big|_{t=0}^{\infty} \\ &= -Q \end{aligned} \tag{1.143}$$

so that $AY + YA^{\mathsf{T}} + Q = 0$ and Y satisfies the same Sylvester equation as X. By uniqueness, we conclude that the integral representation (1.138) holds.

1.9.2 Block Kronecker Products

Let $\mathcal{A}$ now denote a *block* matrix of size $NP \times NP$ with each block having size $P \times P$. We denote the $(i,j)-$th block of $\mathcal{A}$ by the notation A_{ij}; it is a matrix of size $P \times P$. Likewise, let $\mathcal{B}$ denote a second *block* matrix of size $MP \times MP$ with each of its blocks having the same size $P \times P$. We denote the $(i,j)-$th block of $\mathcal{B}$ by the notation B_{ij}; it is a matrix of size $P \times P$:

$$\mathcal{A} = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1N} \\ A_{21} & A_{22} & \dots & A_{2N} \\ \vdots & & & \vdots \\ A_{N1} & A_{N2} & \dots & A_{NN} \end{bmatrix} \tag{1.144a}$$

$$\mathcal{B} = \begin{bmatrix} B_{11} & B_{12} & \dots & B_{1M} \\ B_{21} & B_{22} & \dots & B_{2M} \\ \vdots & & & \vdots \\ B_{M1} & B_{M2} & \dots & B_{MM} \end{bmatrix} \tag{1.144b}$$

The *block* Kronecker product of these two matrices is denoted by $\mathcal{K} = \mathcal{A} \otimes_b \mathcal{B}$ and is defined as the following block matrix of dimensions $NMP^2 \times NMP^2$:

$$\mathcal{K} \triangleq \mathcal{A} \otimes_b \mathcal{B} = \begin{bmatrix} K_{11} & K_{12} & \dots & K_{1N} \\ K_{21} & K_{22} & \dots & K_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ K_{N1} & K_{N2} & \dots & K_{NN} \end{bmatrix} \tag{1.145}$$

where each block K_{ij} is $MP^2 \times MP^2$ and is constructed as follows:

$$K_{ij} = \begin{bmatrix} A_{ij} \otimes B_{11} & A_{ij} \otimes B_{12} & \dots & A_{ij} \otimes B_{1M} \\ A_{ij} \otimes B_{21} & A_{ij} \otimes B_{22} & \dots & A_{ij} \otimes B_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ A_{ij} \otimes B_{M1} & A_{ij} \otimes B_{M2} & \dots & A_{ij} \otimes B_{MM} \end{bmatrix} \tag{1.146}$$

Table 1.2 lists some useful properties of block Kronecker products for matrices $\{\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}\}$ with blocks of size $P \times P$. The last three properties involve the block vectorization operation denoted by *bvec*: it vectorizes each block entry of the matrix and then stacks the resulting columns on top of each other, i.e.,

$$
\text{bvec}(\mathcal{A}) \;\stackrel{\Delta}{=}\; \left[\begin{array}{c} \text{vec}(A_{11}) \\ \vdots \\ \text{vec}(A_{N1}) \\ \hline \text{vec}(A_{21}) \\ \vdots \\ \text{vec}(A_{N2}) \\ \hline \\ \vdots \\ \\ \hline \text{vec}(A_{1N}) \\ \vdots \\ \text{vec}(A_{NN}) \end{array}\right]
\begin{array}{l} \left.\begin{array}{c} \\ \\ \\ \end{array}\right\} \text{first block column of } \mathcal{A} \\ \\ \\ \\ \\ \\ \\ \\ \left.\begin{array}{c} \\ \\ \\ \end{array}\right\} \text{last block column of } \mathcal{A} \end{array}
\tag{1.147}
$$

Expression (1.148) illustrates one of the advantages of working with the bvec operation for block matrices. It compares the effect of the block vectorization operation to that of the regular vec operation. It is seen that bvec preserves the locality of the blocks from the original matrix: entries arising from the same block appear together followed by entries from other blocks. In contrast, in the regular vec construction, entries from different blocks are mixed together.

$$
\left[\begin{array}{c} \circ \\ \circ \\ \square \\ \square \\ \hline \bullet \\ \bullet \\ \blacksquare \\ \blacksquare \\ \hline \triangle \\ \triangle \\ \star \\ \star \\ \hline \blacktriangle \\ \blacktriangle \\ \bigstar \\ \bigstar \end{array}\right]
\stackrel{\text{vec}(\mathcal{A})}{\Longleftarrow}
\underbrace{\left[\begin{array}{cc|cc} \circ & \bullet & \triangle & \blacktriangle \\ \circ & \bullet & \triangle & \blacktriangle \\ \hline \square & \blacksquare & \star & \bigstar \\ \square & \blacksquare & \star & \bigstar \end{array}\right]}_{=\mathcal{A}}
\stackrel{\text{bvec}(\mathcal{A})}{\Longrightarrow}
\left[\begin{array}{c} \circ \\ \circ \\ \bullet \\ \bullet \\ \hline \square \\ \square \\ \blacksquare \\ \blacksquare \\ \hline \triangle \\ \triangle \\ \blacktriangle \\ \blacktriangle \\ \hline \star \\ \star \\ \bigstar \\ \bigstar \end{array}\right]
\tag{1.148}
$$

Table 1.2 Properties for the block Kronecker product (1.145).

	relation	property
1.	$(\mathcal{A}+\mathcal{B})\otimes_b \mathcal{C} = (\mathcal{A}\otimes_b \mathcal{C}) \;+\; (\mathcal{B}\otimes_b \mathcal{C})$	distributive property
2.	$(\mathcal{A}\otimes_b \mathcal{B})(\mathcal{C}\otimes_b \mathcal{D}) = (\mathcal{A}\mathcal{C}\otimes_b \mathcal{B}\mathcal{D})$	multiplication property
3.	$(\mathcal{A}\otimes_b \mathcal{B})^{\mathsf{T}} = \mathcal{A}^{\mathsf{T}}\otimes_b \mathcal{B}^{\mathsf{T}}$	transposition property
4.	$\{\lambda(\mathcal{A}\otimes_b \mathcal{B})\} = \{\lambda_i(\mathcal{A})\lambda_j(\mathcal{B})\}_{i=1,j=1}^{NP,MP}$	eigenvalues
5.	$\mathrm{Tr}(\mathcal{A}\mathcal{B}) = \left(\mathrm{bvec}(\mathcal{B}^{\mathsf{T}})\right)^{\mathsf{T}}\mathrm{bvec}(\mathcal{A})$	trace of matrix product
6.	$\mathrm{bvec}(\mathcal{A}\mathcal{C}\mathcal{B}) = (\mathcal{B}^{\mathsf{T}}\otimes_b \mathcal{A})\mathrm{bvec}(\mathcal{C})$	block vectorization property
7.	$\mathrm{bvec}(xy^{\mathsf{T}}) = y\otimes_b x$	vectorization of outer product

1.10 VECTOR AND MATRIX NORMS

We list in this section several useful vector and matrix norms, which will arise regularly in studies of inference and learning methods.

Definition of norms

If we let X denote an arbitrary real matrix or vector quantity, then a matrix or vector norm, denoted by $\|X\|$, is any function that satisfies the following properties, for any X and Y of compatible dimensions and scalar α:

$$\|X\| \geq 0 \tag{1.149a}$$

$$\|X\| = 0 \text{ if, and only if, } X = 0 \tag{1.149b}$$

$$\|\alpha X\| = |\alpha|\,\|X\| \tag{1.149c}$$

$$\|X+Y\| \leq \|X\| + \|X\|, \quad (\textbf{triangle inequality}) \tag{1.149d}$$

$$\|XY\| \leq \|X\|\,\|Y\|, \quad (\textbf{sub-multiplicative property}) \tag{1.149e}$$

For any column vector $x \in \mathbb{R}^N$ with individual entries $\{x_n\}$, any of the definitions listed in Table 1.3 constitutes a valid vector norm.
It is straightforward to verify the validity of the following inequalities relating these norms:

$$\|x\|_2 \;\leq \|x\|_1 \leq \sqrt{N}\,\|x\|_2 \tag{1.150a}$$

$$\|x\|_\infty \leq \|x\|_2 \leq \sqrt{N}\,\|x\|_\infty \tag{1.150b}$$

$$\|x\|_\infty \leq \|x\|_1 \leq N\,\|x\|_\infty \tag{1.150c}$$

$$\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1 \tag{1.150d}$$

$$\|x\|_\infty \leq \|x\|_p \leq N^{1/p}\,\|x\|_\infty, \quad p \geq 1 \tag{1.150e}$$

There are similarly many useful matrix norms. For any matrix $A \in \mathbb{R}^{N\times M}$ with individual entries $\{a_{\ell k}\}$, any of the definitions listed in Table 1.4 constitutes a valid matrix norm. In particular, the 2−induced norm of A is a special case of the p−induced norm and reduces to the maximum singular value of A — see

Table 1.3 Useful vector norms, where the $\{x_n\}$ denote the entries of the vector x.

vector norm	name
$\|x\|_1 \triangleq \sum_{n=1}^{N} \|x_n\|$	(1−norm or ℓ_1−norm)
$\|x\|_\infty \triangleq \max_{1\le n\le N} \|x_n\|$	(∞−norm or ℓ_∞−norm)
$\|x\|_2 \triangleq \left(\sum_{n=1}^{N} \|x_n\|^2\right)^{1/2}$	(Euclidean or ℓ_2−norm, also written as $\|x\|$)
$\|x\|_p \triangleq \left(\sum_{n=1}^{N} \|x_n\|^p\right)^{1/p}$	(p−norm or ℓ_p−norm, for any real $p \ge 1$)

next Example 1.13. A fundamental result in matrix theory is that all matrix norms in finite dimensional spaces are *equivalent*. Specifically, if $\|A\|_a$ and $\|A\|_b$ denote two generic matrix norms, then there exist positive constants c_ℓ and c_u that bound one norm by the other from above and from below, namely,

$$c_\ell \, \|A\|_b \;\le\; \|A\|_a \;\le\; c_u \, \|A\|_b \tag{1.151}$$

The values of $\{c_\ell, c_u\}$ are independent of the matrix entries but they may be dependent on the matrix dimensions. Vector norms are also equivalent to each other.

Table 1.4 Useful matrix norms, where the $\{a_{nm}\}$ denote the entries of A.

matrix norm	name
$\|A\|_1 \triangleq \max_{1\le m\le M} \left(\sum_{n=1}^{N} \|a_{nm}\|\right)$	(1−norm, or maximum absolute column sum)
$\|A\|_\infty \triangleq \max_{1\le n\le N} \left(\sum_{m=1}^{M} \|a_{nm}\|\right)$	(∞−norm, or maximum absolute row sum)
$\|A\|_{\rm F} \triangleq \sqrt{\mathrm{Tr}(A^\mathsf{T} A)}$	(Frobenius norm)
$\|A\|_p \triangleq \max_{x\ne 0} \dfrac{\|Ax\|_p}{\|x\|_p}$	(p−induced norm for any real $p \ge 1$)
$\|A\|_2 \triangleq \max_{x\ne 0} \dfrac{\|Ax\|}{\|x\|}$	(2−induced norm)

Example 1.13 **(Spectral norm of a matrix)** Assume $N \ge M$ and consider the $M \times M$ square matrix $A^\mathsf{T} A$. Using the Rayleigh-Ritz characterization (1.17b) for the maximum

eigenvalue of a matrix we have that

$$\lambda_{\max}(A^{\mathsf T}A) = \max_{x\neq 0}\left(\frac{x^{\mathsf T}A^{\mathsf T}Ax}{x^{\mathsf T}x}\right) \;=\; \max_{x\neq 0}\left(\frac{\|Ax\|^2}{\|x\|^2}\right) \tag{1.152}$$

But we already know from the argument in Appendix 1.B that $\sigma_1^2 = \lambda_{\max}(A^{\mathsf T}A)$. We conclude that the largest singular value of A satisfies:

$$\sigma_1 \;=\; \max_{x\neq 0}\left(\frac{\|Ax\|}{\|x\|}\right) \tag{1.153}$$

This maximum value is achieved if we select $x = v_1$ (i.e., as the right singular vector corresponding to σ_1). Indeed, using $Av_1 = \sigma_1 u_1$ we get

$$\frac{\|Av_1\|^2}{\|v_1\|^2} \;=\; \frac{\|\sigma_1 u_1\|^2}{\|v_1\|^2} \;=\; \sigma_1^2 \tag{1.154}$$

since $\|u_1\| = \|v_1\| = 1$. We therefore find that the square of the maximum singular value, σ_1^2, measures the maximum energy gain from x to Ax. The same conclusion holds when $N \le M$ since then $\sigma_1^2 = \lambda_{\max}(AA^{\mathsf T})$ and the argument can be repeated to conclude that

$$\sigma_1 \;=\; \max_{x\neq 0}\left(\frac{\|A^{\mathsf T}x\|}{\|x\|}\right) \tag{1.155}$$

The maximum singular value of a matrix is called its *spectral norm* or its 2-induced norm, also written as

$$\sigma_1 = \|A\|_2 = \|A^{\mathsf T}\|_2 = \max_{\|x\|=1}\|Ax\| \;=\; \max_{\|x\|=1}\|A^{\mathsf T}x\| \tag{1.156}$$

Dual norms

An important concept in matrix theory is that of the *dual norm*. Let $\|\cdot\|$ denote some vector norm in $\mathbb{R}^M$. The associated dual norm is denoted by $\|x\|_\star$ and defined as

$$\|x\|_\star \;\stackrel{\Delta}{=}\; \sup_y \left\{x^{\mathsf T}y \mid \|y\| \le 1\right\} \tag{1.157}$$

In other words, we consider all vectors y that lie inside the ball $\|y\| \le 1$ and examine their transformation by x through the inner product $x^{\mathsf T}y$. The largest value this transformation attains is taken as the dual norm of x. It is shown in Prob. 1.24 that $\|x\|_\star$ is a valid vector norm and that it can be expressed equivalently as

$$\|x\|_\star \;\stackrel{\Delta}{=}\; \sup_{y\neq 0}\left\{\frac{x^{\mathsf T}y}{\|y\|}\right\} \tag{1.158}$$

where we scale the inner product by $\|y\|$ and compute the supremum over all nonzero vectors y. Using (1.158), we can readily write the following Cauchy-Schwarz inequality involving a norm and its dual:

$$\boxed{x^{\mathsf T}y \;\le\; \|y\|\,\|x\|_\star} \tag{1.159}$$

This result will be useful in our analysis of inference methods. This is because we will often deal with norms other than the Euclidean norm, and we will need to bound inner products between vectors. The above inequality shows that the bound will involve the product of two norms: the regular norm and its dual.

For the sake of illustration, let us determine the dual norm of the Euclidean norm, $\|\cdot\|_2$. To do so, we need to assess:

$$\|x\|_\star = \sup_y \left\{ x^\mathsf{T} y \mid \|y\|_2 \leq 1 \right\} \tag{1.160}$$

We know from the classical Cauchy-Schwarz inequality for inner products in Euclidean space that $x^\mathsf{T} y \leq \|x\|_2 \, \|y\|_2$, with equality when the vectors are parallel to each other and pointing in the same direction, say, $y = \alpha x$ for some $\alpha > 0$. Given that the norm of y should be bounded by one, we must set $\alpha \leq 1/\|x\|_2$. The inner product $x^\mathsf{T} y$ becomes $\alpha\|x\|_2^2$ and it is maximized when $\alpha = 1/\|x\|$, in which case we conclude that $\|x\|_\star = \|x\|_2$. In other words, the dual norm of the Euclidean norm is the Euclidean norm itself.

Several other dual norms are determined in Probs. 1.26–1.29 for both vector and matrix norms, where the definition of the dual norm for matrices is taken as:

$$\|A\|_\star \;\stackrel{\Delta}{=}\; \sup_Y \left\{ \mathrm{Tr}(A^\mathsf{T} Y) \,|\, \|Y\| \leq 1 \right\} \tag{1.161}$$

We collect the results into Table 1.5. The last line introduces the nuclear norm of a matrix, which is equal to the sum of its singular values. The notation for the nuclear norm continues to use the star subscript. The results in the table show that we can interpret the $(\ell_1, \ell_2, \ell_p, \ell_\infty)-$norms of vectors in the equivalent forms:

$$\|x\|_2 = \sup_y \left\{ x^\mathsf{T} y \mid \|y\|_2 \leq 1 \right\} \tag{1.162a}$$

$$\|x\|_\infty = \sup_y \left\{ x^\mathsf{T} y \mid \|y\|_1 \leq 1 \right\} \tag{1.162b}$$

$$\|x\|_1 = \sup_y \left\{ x^\mathsf{T} y \mid \|y\|_\infty \leq 1 \right\} \tag{1.162c}$$

$$\|x\|_q = \sup_y \left\{ x^\mathsf{T} y \mid \|y\|_p \leq 1 \right\}, \quad p, q \geq 1, \quad 1/p + 1/q = 1 \tag{1.162d}$$

Likewise, the Forbenius and nuclear norms of a matrix A can be interpreted as corresponding to:

$$\|A\|_\mathrm{F} = \sup_Y \left\{ \mathrm{Tr}(A^\mathsf{T} Y) \,|\, \|Y\|_\mathrm{F} \leq 1 \right\} \tag{1.163a}$$

$$\|A\|_\star = \sup_Y \left\{ \mathrm{Tr}(A^\mathsf{T} Y) \,|\, \|Y\|_2 \leq 1 \right\} \tag{1.163b}$$

It is straightforward to verify the validity of the following inequalities relating

Table 1.5 List of dual norms for vectors and matrices.

original norm	dual norm
$\|x\|_2$	$\|x\|_2$
$\|x\|_1$	$\|x\|_\infty$
$\|x\|_\infty$	$\|x\|_1$
$\|x\|_p$	$\|x\|_q,\ p,q \geq 1,\ 1/p + 1/q = 1$
$\|A\|_{\rm F}$	$\|A\|_{\rm F}$
$\|A\|_2$	$\|A\|_\star = \sum_{r=1}^{\text{rank}(A)} \sigma_r,$ (nuclear norm)

several matrix norms for matrices A of size $N \times M$ and rank r:

$$M^{-1/2}\,\|A\|_\infty \leq \|A\|_2 \leq N^{1/2}\,\|A\|_\infty \tag{1.164a}$$
$$N^{-1/2}\,\|A\|_1 \leq \|A\|_2 \leq M^{1/2}\,\|A\|_1 \tag{1.164b}$$
$$\|A\|_\infty \leq M^{1/2}\,\|A\|_2 \leq M\,\|A\|_1 \tag{1.164c}$$
$$\|A\|_2 \leq \|A\|_{\rm F} \leq \sqrt{r}\,\|A\|_2 \tag{1.164d}$$
$$\|A\|_{\rm F} \leq \|A\|_\star \leq \sqrt{r}\,\|A\|_{\rm F} \tag{1.164e}$$
$$\|A\|_2^2 \leq \|A\|_1 \times \|A\|_\infty \tag{1.164f}$$

ρ-norm[1]

In this section and the next we introduce two specific norms; the discussion can be skipped on a first reading. Let B denote an $N \times N$ real matrix with eigenvalues $\{\lambda_n\}$. The spectral radius of B, denoted by $\rho(B)$, is defined as

$$\boxed{\rho(B) \;\stackrel{\Delta}{=}\; \max_{1\leq n\leq N}\; |\lambda_n|} \tag{1.165}$$

A fundamental result in matrix theory asserts that every matrix admits a so-called *canonical Jordan decomposition*, which is of the form

$$B = UJU^{-1} \tag{1.166}$$

for some invertible matrix U and where

$$J = \text{blkdiag}\{J_1, J_2, \ldots, J_R\} \tag{1.167}$$

is a block diagonal matrix, say with R blocks. When B happens to be symmetric, we already know from the spectral decomposition (1.15a) that J will be diagonal and U will be orthogonal so that $U^{-1} = U^{\mathsf{T}}$. More generally, each block J_r will have a Jordan structure with an eigenvalue λ_r on its diagonal entries, unit entries on the first sub-diagonal, and zeros everywhere else. For example, for a block of

[1] The two sections on the block-maximum and $\rho-$norms can be skipped on a first reading.

size 4×4:

$$J_r = \begin{bmatrix} \lambda_r & & & \\ 1 & \lambda_r & & \\ & 1 & \lambda_r & \\ & & 1 & \lambda_r \end{bmatrix} \tag{1.168}$$

Let ϵ denote an arbitrary positive scalar that we are free to choose and define the $N \times N$ diagonal scaling matrix:

$$D \triangleq \text{diag}\left\{\epsilon, \epsilon^2, \ldots, \epsilon^N\right\} \tag{1.169}$$

We can use the matrix U originating from B to define the following matrix norm, denoted by $\|\cdot\|_\rho$, for any matrix A of size $N \times N$:

$$\boxed{\|A\|_\rho \triangleq \left\|DU^{-1}AUD^{-1}\right\|_1} \tag{1.170}$$

in terms of the $1-$norm (i.e., maximum absolute column sum) of the matrix product on the right-hand side. It is not difficult to verify that the transformation (1.170) is a valid matrix norm, namely, that it satisfies the following properties, for any matrices A and C of compatible dimensions and for any scalar α:

(a) $\|A\|_\rho \geq 0$ with $\|A\|_\rho = 0$ if, and only if, $A = 0$
(b) $\|\alpha A\|_\rho = |\alpha| \, \|A\|_\rho$
(c) $\|A + C\|_\rho \leq \|A\|_\rho + \|C\|_\rho$, **(triangular inequality)**
(d) $\|AC\|_\rho \leq \|A\|_\rho \, \|C\|_\rho$, **(sub-multiplicative property)** (1.171)

One important property of the $\rho-$norm defined by (1.170) is that when it is applied to the matrix B itself, it will hold that:

$$\rho(B) \leq \|B\|_\rho \leq \rho(B) + \epsilon \tag{1.172}$$

That is, the $\rho-$norm of B lies between two bounds defined by its spectral radius. It follows that if the matrix B happens to be stable to begin with, so that $\rho(B) < 1$, then we can always select ϵ small enough to ensure $\|B\|_\rho < 1$.

The matrix norm defined by (1.170) is an induced norm relative to the following vector norm:

$$\|x\|_\rho \triangleq \|DU^{-1}x\|_1 \tag{1.173}$$

That is, for any matrix A, it holds that

$$\|A\|_\rho = \max_{x \neq 0} \left\{ \frac{\|Ax\|_\rho}{\|x\|_\rho} \right\} \tag{1.174}$$

Proof of (1.174): Indeed, using (1.173), we first note that for any vector $x \neq 0$:

$$
\begin{aligned}
\|Ax\|_\rho &= \|DU^{-1}Ax\|_1 \\
&= \|DU^{-1}AUD^{-1}DU^{-1}x\|_1 \\
&\leq \|DU^{-1}AUD^{-1}\|_1 \, \|DU^{-1}x\|_1 \\
&= \|A\|_\rho \, \|x\|_\rho
\end{aligned} \tag{1.175}
$$

so that

$$
\max_{x \neq 0} \left\{ \frac{\|Ax\|_\rho}{\|x\|_\rho} \right\} \leq \|A\|_\rho \tag{1.176}
$$

To show that equality holds in (1.176), it is sufficient to exhibit one nonzero vector x_o that attains it. Let k_o denote the index of the column that attains the maximum absolute column sum in the matrix product $DU^{-1}AUD^{-1}$. Let e_{k_o} denote the column basis vector of size $N \times 1$ with one at location k_o and zeros elsewhere. Select

$$
x_o \triangleq UD^{-1}e_{k_o} \tag{1.177}
$$

Then, it is straightforward to verify that

$$
\|x_o\|_\rho \triangleq \|DU^{-1}x_o\|_1 \overset{(1.177)}{=} \|e_{k_o}\|_1 = 1 \tag{1.178}
$$

and

$$
\begin{aligned}
\|Ax_o\|_\rho &\triangleq \|DU^{-1}Ax_o\|_1 \\
&= \|DU^{-1}AUD^{-1}DU^{-1}x_o\|_1 \\
&\overset{(1.177)}{=} \|DU^{-1}AUD^{-1}e_{k_o}\|_1 \\
&\overset{(1.170)}{=} \|A\|_\rho
\end{aligned} \tag{1.179}
$$

so that, for this particular vector, we have

$$
\frac{\|Ax_o\|_\rho}{\|x_o\|_\rho} = \|A\|_\rho \tag{1.180}
$$

as desired.

■

Block maximum norm

Let

$$
x \triangleq \text{blkcol}\{x_1, x_2, \ldots, x_N\} \tag{1.181}
$$

denote an $N \times 1$ *block* column vector whose individual entries $\{x_k\}$ are themselves vectors of size $M \times 1$ each. The block maximum norm of x is denoted by $\|x\|_{b,\infty}$ and is defined as

$$
\|x\|_{b,\infty} \triangleq \max_{1 \leq k \leq N} \|x_k\| \tag{1.182}
$$

That is, $\|x\|_{b,\infty}$ is equal to the largest Euclidean norm of its block components. This vector norm induces a block maximum matrix norm. Let $\mathcal{A}$ denote an arbitrary $N \times N$ block matrix with individual block entries of size $M \times M$ each. Then, the block maximum norm of $\mathcal{A}$ is defined as

$$\|\mathcal{A}\|_{b,\infty} \;\triangleq\; \max_{x\neq 0}\left\{\frac{\|\mathcal{A}x\|_{b,\infty}}{\|x\|_{b,\infty}}\right\} \tag{1.183}$$

The block maximum norm has several useful properties — see Prob. 1.23:

(a) Let $\mathcal{U} = \text{diag}\{U_1, U_2, \ldots, U_N\}$ denote an $N \times N$ block diagonal matrix with $M \times M$ orthogonal blocks $\{U_k\}$. Then, transformations by $\mathcal{U}$ do not modify the block maximum norm, i.e., it holds that $\|\mathcal{U}x\|_{b,\infty} = \|x\|_{b,\infty}$ and $\|\mathcal{U}\mathcal{A}\mathcal{U}^{\mathsf{T}}\|_{b,\infty} = \|\mathcal{A}\|_{b,\infty}$.

(b) Let $\mathcal{D} = \text{diag}\{D_1,\ D_2,\ \ldots, D_N\}$ denote an $N \times N$ block diagonal matrix with $M \times M$ symmetric blocks $\{D_k\}$. Then, $\rho(\mathcal{D}) = \|\mathcal{D}\|_{b,\infty}$.

(c) Let A be an $N \times N$ matrix and define $\mathcal{A} = A \otimes I_M$ whose blocks are therefore of size $M \times M$ each. If A is left-stochastic (i.e., the entries on each column of A add up to one, as defined further ahead by (1.193)), then $\|\mathcal{A}^{\mathsf{T}}\|_{b,\infty} = 1$.

(d) For any block diagonal matrix $\mathcal{D}$, and any left-stochastic matrices $\mathcal{A}_1 = A_1 \otimes I_M$ and $\mathcal{A}_2 = A_2 \otimes I_M$ constructed as in part (c), it holds that

$$\rho\left(\mathcal{A}_2^{\mathsf{T}}\,\mathcal{D}\,\mathcal{A}_1^{\mathsf{T}}\right) \;\leq\; \|\mathcal{D}\|_{b,\infty} \tag{1.184}$$

(e) If the matrix $\mathcal{D}$ in part (d) has symmetric blocks, it holds that

$$\rho\left(\mathcal{A}_2^{\mathsf{T}}\,\mathcal{D}\,\mathcal{A}_1^{\mathsf{T}}\right) \;\leq\; \rho(\mathcal{D}) \tag{1.185}$$

1.11 PERTURBATION BOUNDS ON EIGENVALUES[2]

We state below two useful results that bound matrix eigenvalues.

Weyl theorem

The first result, known as *Weyl theorem*, shows how the eigenvalues of a symmetric matrix are disturbed through additive perturbations to the entries of the matrix. Thus, let $\{A', A, \Delta A\}$ denote arbitrary $N \times N$ real symmetric matrices with ordered eigenvalues $\{\lambda_m(A'), \lambda_m(A), \lambda_m(\Delta A)\}$, i.e.,

$$\lambda_1(A) \;\geq\; \lambda_2(A) \;\geq\; \ldots \;\geq\; \lambda_N(A) \tag{1.186}$$

and similarly for the eigenvalues of $\{A', \Delta A\}$, with the subscripts 1 and N representing the largest and smallest eigenvalues, respectively. Weyl Theorem states that if A is perturbed to

$$A' \;=\; A \;+\; \Delta A \tag{1.187}$$

[2] This section can be skipped on a first reading.

then the eigenvalues of the new matrix are bounded as follows:

$$\boxed{\lambda_n(A) \;+\; \lambda_N(\Delta A) \;\leq\; \lambda_n(A') \;\leq\; \lambda_n(A) \;+\; \lambda_1(\Delta A)} \tag{1.188}$$

for $1 \leq n \leq N$. In particular, it follows that the maximum eigenvalue is perturbed as follows:

$$\lambda_{\max}(A) \;+\; \lambda_{\min}(\Delta A) \;\leq\; \lambda_{\max}(A') \;\leq\; \lambda_{\max}(A) \;+\; \lambda_{\max}(\Delta A) \tag{1.189}$$

In the special case when $\Delta A \geq 0$, we conclude from (1.188) that $\lambda_n(A') \geq \lambda_n(A)$ for all $n = 1, 2, \ldots, N$.

Gershgorin theorem

The second result, known as *Gershgorin theorem*, specifies circular regions within which the eigenvalues of a matrix are located. Thus, consider an $N \times N$ real matrix A with scalar entries $\{a_{\ell k}\}$. With each diagonal entry $a_{\ell\ell}$ we associate a disc in the complex plane centered at $a_{\ell\ell}$ and with radius

$$r_\ell \;\stackrel{\Delta}{=}\; \sum_{k\neq\ell,k=1}^{N} |a_{\ell k}| \tag{1.190}$$

That is, r_ℓ is equal to the sum of the magnitudes of the non-diagonal entries on the same row as $a_{\ell\ell}$. We denote the disc by D_ℓ; it consists of all points that satisfy

$$D_\ell \;=\; \left\{ z \in \mathbb{C}^N \text{ such that } |z - a_{\ell\ell}| \leq r_\ell \right\} \tag{1.191}$$

The theorem states that the spectrum of A (i.e., the set of all its eigenvalues, denoted by $\lambda(A)$) is contained in the union of all N Gershgorin discs:

$$\lambda(A) \subset \left\{ \bigcup_{\ell=1}^{N} D_\ell \right\} \tag{1.192}$$

A stronger statement of the Gershgorin theorem covers the situation in which some of the Gershgorin discs happen to be disjoint. Specifically, if the union of L of the discs is disjoint from the union of the remaining $N - L$ discs, then the theorem further asserts that L eigenvalues of A will lie in the first union of L discs and the remaining $N - L$ eigenvalues of A will lie in the second union of $N - L$ discs.

1.12 Stochastic Matrices

Consider $N \times N$ matrices A with nonnegative entries, $\{a_{\ell k} \geq 0\}$. The matrix $A = [a_{\ell k}]$ is said to be left-stochastic if it satisfies

$$A^{\mathsf{T}}\mathbb{1} = \mathbb{1}, \qquad\qquad (\textbf{left-stochastic}) \tag{1.193}$$

where $\mathbb{1}$ denotes the column vector whose entries are all equal to one. That is, the entries on each column of A should add up to one. The matrix A is said to be right-stochastic if

$$A\mathbb{1} = \mathbb{1}, \qquad (\textbf{right-stochastic}) \tag{1.194}$$

so that the entries on each row of A add up to one. The matrix A is doubly-stochastic if the entries on each of its columns and on each of its rows add up to one (i.e., if it is both left and right stochastic):

$$A\mathbb{1} = \mathbb{1}, \quad A^{\mathsf{T}}\mathbb{1} = \mathbb{1}, \qquad (\textbf{doubly-stochastic}) \tag{1.195}$$

Stochastic matrices arise frequently in the study of Markov chains, multi-agent networks, and signals over graphs. The following statement lists two properties of stochastic matrices:

(a) The spectral radius of A is equal to one, $\rho(A) = 1$. It follows that all eigenvalues of A lie inside the unit disc, i.e., $|\lambda(A)| \leq 1$. The matrix A may have multiple eigenvalues with magnitude equal to one — see Prob. 1.49.

(b) Assume A is additionally a primitive matrix, i.e., there exists some finite integer power of A such that all its entries are strictly positive:

$$[A^{n_o}]_{\ell k} \; > \; 0, \quad \text{for some integer } n_o > 0 \tag{1.196}$$

and for all $1 \leq \ell, k \leq N$. Then, the matrix A will have a single eigenvalue at one (i.e., the eigenvalue at one will have multiplicity one). All other eigenvalues of A will lie strictly inside the unit circle. Moreover, with proper sign scaling, all entries of the right-eigenvector of A corresponding to the single eigenvalue at one will be strictly positive, namely, if we let p denote this right-eigenvector with entries $\{p_k\}$ and normalize its entries to add up to one, then

$$Ap = p, \quad \mathbb{1}^{\mathsf{T}}p = 1, \quad p_k > 0, \quad k = 1, 2, \ldots, N \tag{1.197}$$

We refer to p as the *Perron eigenvector* of A. All other eigenvectors of A associated with the other eigenvalues will have at least one negative or complex entry.

1.13 COMPLEX-VALUED MATRICES

Although the presentation in the earlier sections has focused exclusively on real-valued matrices, most of the concepts and results extend almost effortlessly to complex-valued matrices. For example, in relation to symmetry, if $A \in \mathbb{C}^{N\times N}$, then the matrix A will be said to be *Hermitian* if it satisfies

$$A = A^*, \quad (\textbf{Hermitian symmetry}) \tag{1.198}$$

where the symbol $*$ denotes complex conjugate transposition — recall (1.5). This notion extends the definition of matrix symmetry (which requires $A = A^{\mathsf{T}}$) to the complex case. Hermitian matrices can again be shown to have only real eigenvalues and their spectral decomposition will now take the form:

$$A = U\Lambda U^* \tag{1.199}$$

where Λ continues to be a diagonal matrix with the eigenvalues of A, while U is now a *unitary* (as opposed to an orthogonal) matrix, namely, it satisfies

$$UU^* = U^*U = I_N \tag{1.200}$$

In other words, most of the results discussed so far extend rather immediately by replacing transposition by complex conjugation, such as replacing x^{T} by x^* and A^{T} by A^*. For example, while the squared Euclidean norm of a vector $x \in \mathbb{R}^N$ is given by $\|x\|^2 = x^{\mathsf{T}}x$, the same squared norm for a vector $x \in \mathbb{C}^N$ will be given by $\|x\|^2 = x^*x$. In this way, the Rayleigh-Ritz characterization of the smallest and largest eigenvalues of A will become

$$\lambda_{\min} = \min_{\|x\|=1} \left\{x^*Ax\right\}, \qquad \lambda_{\max} = \max_{\|x\|=1} \left\{x^*Ax\right\} \tag{1.201}$$

Likewise, positive-definite matrices will be ones that satisfy

$$v^*Av > 0, \;\; \text{for any } v \neq 0 \in \mathbb{C}^N \tag{1.202}$$

These matrices will continue to have positive eigenvalues and positive determinants. In addition, the range space and nullspace of A will be defined similarly to the real case:

$$\mathcal{R}(A) \;\stackrel{\Delta}{=}\; \left\{q \in \mathbb{C}^N \mid \text{such that } q = Ap \text{ for some } p \in \mathbb{C}^M\right\} \tag{1.203a}$$

$$\mathcal{N}(A) \;\stackrel{\Delta}{=}\; \left\{p \in \mathbb{C}^M \mid \text{such that } Ap = 0\right\} \tag{1.203b}$$

with the properties

$$z \in \mathcal{N}(A^*), \;\; q \in \mathcal{R}(A) \quad \Longrightarrow \quad z^*q = 0 \tag{1.204a}$$

$$\mathcal{R}(A^*) = \mathcal{R}(A^*A) \tag{1.204b}$$

$$\mathcal{N}(A) = \mathcal{N}(A^*A) \tag{1.204c}$$

Moreover, two square matrices A and B will be congruent if $A = QBQ^*$, for some nonsingular Q. Finally, the SVD of $A \in \mathbb{C}^{N\times M}$ will now take the following form with unitary matrices $U \in \mathbb{C}^{N\times N}$ and $V \in \mathbb{C}^{M\times M}$:

(a) If $N \leq M$, then Σ is $N \times N$ and

$$A = U \begin{bmatrix} \Sigma & 0 \end{bmatrix} V^* \tag{1.205}$$

(b) If $N \geq M$, then Σ is $M \times M$ and

$$A = U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^* \tag{1.206}$$

1.14 COMMENTARIES AND DISCUSSION

Linear algebra and matrix theory. The presentation in this chapter follows the overviews from Sayed (2003,2008,2014a). Throughout our treatment of inference and learning theories, the reader will be exposed to a variety of concepts from linear algebra and matrix theory in a motivated manner. In this way, after progressing sufficiently enough into our treatment, readers will be able to master many useful concepts. Several of these concepts are summarized in this chapter. If additional help is needed, some accessible references on matrix theory are the works by MacDuffee (1946), Gantmacher (1959), Bellman (1970), Horn and Johnson (1990), Golub and Van Loan (1996), Meyer (2001), Laub (2004), and Bernstein (2018) Accessible references on linear algebra are the books by Halmos (1974), Strang (1988,2009), Gelfand (1989), Lay (1994), Lax (1997), Lay, Lay, and McDonald (2014), Hogben (2014), and Nicholson (2019).

Kronecker products. We introduced the Kronecker product in Sec. 1.9. This product provides a useful and compact representation for generating a block matrix structure from two separate matrices. We illustrated two useful applications of Kronecker products in Examples 1.11 and 1.12 in the context of Lyapunov equations. There are of course many other applications. The notion of Kronecker products was introduced by the German mathematician **Johann Zehfuss (1832–1901)** in the work by Zehfuss (1858) — see the historical accounts by Henderson, Pukelsheim, and Searle (1983) and Hackbusch (2012). In his article, Zehfuss (1858) introduced the determinantal formula

$$\det(A \otimes B) = (\det(A))^M \, (\det(B))^N \tag{1.207}$$

for square matrices A and B of sizes $N \times N$ and $M \times M$, respectively. This formula was later attributed erroneously by Hensel (1891) to the German mathematician **Leopold Kronecker (1823–1891)** who discussed it in some of his lectures in the 1880's. The operation $\otimes$ became subsequently known as the Kronecker product instead of the more appropriate "Zehfuss product". Useful surveys on Kronecker products, their properties and applications appear in Henderson and Searle (1981b), Regalia and Mitra (1989), and Van Loan (2000). Useful references on block Kronecker products are the works by Tracy and Singh (1972), Koning, Neudecker, and Wansbeek (1991), and Liu (1999). The block Kronecker product (1.145) is also known as the Tracy-Singh product.

Schur complements. According to the historical overview by Puntanen and Styan (2005), the designation "Schur complement" is due to Haynsworth (1968) in her study of the inertia of a block-partitioned matrix. If we consider a block symmetric matrix of the form

$$S = \begin{bmatrix} A & B \\ B^{\mathsf{T}} & D \end{bmatrix}, \quad A = A^{\mathsf{T}}, \; D = D^{\mathsf{T}} \tag{1.208}$$

with a nonsingular A, we recognize that the following block triangular factorization of S amounts to a congruence relation:

$$S = \begin{bmatrix} I & 0 \\ B^{\mathsf{T}}A^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & \Delta_A \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix} \tag{1.209}$$

where $\Delta_A = D - B^{\mathsf{T}}A^{-1}B$. It follows that

$$\text{In}(S) \;=\; \text{In}(A) \;+\; \text{In}(\Delta_A) \tag{1.210}$$

where the addition operation means that the individual components of the inertia measure, namely, I_+, I_-, I_0, are added together. This inertia additivity formula was derived by Haynsworth (1968) and the factor Δ_A was referred to as the "Schur complement" relative to A. The attribution to "Schur" is because the determinantal formula (1.211) involving Δ_A was given by the German mathematician **Issai Schur (1875–1941)** in a famous paper by Schur (1917), where he studied the characterization of functions that are analytic and contractive inside the unit disc — see the the survey articles by Cottle (1974), Kailath (1986), and Kailath and Sayed (1995), and the books by Constantinescu (1996) and Zhang (2005). In Schur (1917), the following expression appeared, which is now easy to conclude from the block factorization expression (1.209):

$$\det(S) = \det(A)\,\det(\Delta_A) \tag{1.211}$$

Schur was a student of the German mathematician **Georg Frobenius (1849–1917)**. His determinantal formula extended a special case studied by Frobenius (1908), which corresponds to the situation in which D is a scalar δ, and B is a column vector b, i.e.,

$$\det\left(\begin{bmatrix} A & b \\ b^{\mathsf{T}} & \delta \end{bmatrix}\right) = \det(A)\,(\delta - b^{\mathsf{T}}A^{-1}b) \tag{1.212}$$

The fact that congruence preserves inertia, as illustrated in Example 1.5 in connection with Schur complements, was established by Sylvester (1852) and is known as the Sylvester law of inertia.

The block triangular factorization formulas (1.63)–(1.65), written in various equivalent forms, appear less directly in the work of Schur (1917) and more explicitly in the works by Banachiewicz (1937a,b), Aitken (1939), Hotelling (1943a,b), and Duncan (1944) — see the overviews by Henderson and Searle (1981a) and Puntanen, Styan, and Isotalo (2011). The matrix inversion formula (1.81) was apparently first given by Duncan (1944) and Guttman (1946), though it is often attributed to Sherman and Morrison (1949,1950) and Woodbury (1950) and referred to as the Sherman-Morrison-Woodbury formula. A special case when $C = 1$, B is a column vector b, and D is a row vector, d^{T}, appeared in Bartlett (1951), namely,

$$(A + bd^{\mathsf{T}})^{-1} = A^{-1} - \frac{A^{-1}bd^{\mathsf{T}}A^{-1}}{1 + d^{\mathsf{T}}A^{-1}b} \tag{1.213}$$

This particular inversion formula was also introduced independently by Plackett (1950) in his study of recursive updates of least-squares problems, as we are going to discuss later in Sec. 50.3.2. Useful overviews on the origin of the matrix inversion formula appear in Householder (1953,1957,1964), Henderson and Searle (1981a), and Hager (1989). One of the first uses of the formula in the context of filtering theory appears to be Kailath (1960) and Ho (1963).

Spectral theorem. We established the spectral theorem for finite-dimensional matrices in Sec. 1.1, which states that every $N \times N$ symmetric matrix A has N real eigenvalues, $\{\lambda_n\}$, and a complete set of N orthonormal eigenvectors, $\{u_n\}$. This means that A can be expressed in either form:

$$A \;=\; U\Lambda U^{\mathsf{T}} \;=\; \sum_{n=1}^{N} \lambda_n u_n u_n^{\mathsf{T}} \tag{1.214}$$

where U is an orthogonal matrix whose columns are the $\{u_n\}$ and Λ is a diagonal matrix with real entries $\{\lambda_n\}$. For any vector $x \in \mathbb{R}^n$, we introduce the change of variables $y = U^{\mathsf{T}}x$ and let $\{y_n\}$ denote the individual entries of y. Then, one useful consequence of the spectral decomposition (1.214) is that every generic quadratic term of the form $x^{\mathsf{T}}Ax$ can be decomposed into the sum of N elementary quadratic terms, namely,

$$x^{\mathsf{T}}Ax = (x^{\mathsf{T}}U)\Lambda(U^{\mathsf{T}}x) \;=\; y^{\mathsf{T}}\Lambda y \;=\; \sum_{n=1}^{N} \lambda_n y_n^2 \tag{1.215}$$

The property that every symmetric matrix is diagonalizable holds more generally and is part of the spectral theory of self-adjoint linear operators in Hilbert space (which extends the study of symmetric finite-dimensional matrix mappings to infinite dimensional mappings defined over spaces endowed with inner products) — an accessible overview is given by Halmos (1963,1974,2013). It is sufficient for our purposes to focus on finite-dimensional matrices. The result in this case was first established by the French mathematician **Augustine Cauchy (1789–1857)** in the work by Cauchy (1829) — see the useful historical account by Hawkins (1975). It was later generalized to the operator setting by the Hungarian-American mathematician **John von Neumann (1903–1957)** in the work by von Neumann (1929,1932) in his studies of linear operators in the context of quantum mechanics — see the overview in the volume edited by Bródy and Vámos (1995). The derivation of the spectral theorem in Appendix 1.A relies on the fundamental theorem of algebra, which guarantees that every polynomial of order N has N roots — see, e.g., the text by Fine and Rosenberger (1997) and Carrera (1992). The argument in the appendix is motivated by the presentations in Horn and Johnson (1990), Trefethen and Bau (1997), Calafiore and El Ghaoui (2014), and Nicholson (2019).

QR decomposition. The matrix decomposition (1.102) is a restatement of the Gram-Schmidt orthonormalization procedure, whereby the columns of a matrix A are replaced by an orthonormal basis for the linear subspace that is spanned by them. The Gram-Schmidt procedure is named after the Danish and German mathematicians **Jorgen Gram (1850–1916)** and **Erhard Schmidt (1876–1959)**, respectively. The latter published the procedure in the work by Schmidt (1907,1908) and acknowledged that his algorithm is the same as one published earlier by Gram (1883). Nevertheless, a similar construction was already proposed over half a century before by the French mathematician **Pierre-Simon Laplace (1749–1827)** in the treatise by Laplace (1812). He orthogonalized the columns of a (tall) observation matrix $A \in \mathbb{R}^{N\times M}$ in order to solve a least-squares problem of the form:

$$\min_{w\in\mathbb{R}^M} \|y - Aw\|^2 \tag{1.216}$$

His solution method is a precursor of the QR-method for solving least-squares problems. Specifically, assuming A has full rank, we introduce the full QR decomposition of A,

$$A = Q \begin{bmatrix} R \\ 0 \end{bmatrix} \tag{1.217}$$

where Q is $N \times N$ orthogonal and R is $M \times M$ upper-triangular with positive diagonal entries. We further let

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \triangleq Q^\mathsf{T} y \tag{1.218}$$

where z_1 is $M \times 1$. It then follows that

$$\|y - Aw\|^2 = \|z_1 - Rw\|^2 + \|z_2\|^2 \tag{1.219}$$

so that the solution to (1.216) is obtained by solving the triangular system of equations $R\widehat{w} = z_1$, i.e.,

$$\widehat{w} \;=\; R^{-1}z_1 \tag{1.220}$$

According to Bjorck (1996), the earliest work linking the names of Gram and Schmidt to the orthogonalization procedure appears to be Wong (1935). Today, the QR decomposition is one of the main tools in modern numerical analysis. For further information and discussions, the reader may refer to the texts by Bjorck (1996), Golub and Van Loan (1996), and Trefethen and Bau (1997).

Singular value decomposition. The singular value decomposition (SVD) is one of the most powerful matrix decompositions, valid for both rectangular and square matrices and useful for both analysis and numerical computations. It has a rich history with contributions from notable mathematicians. The article by Stewart (1993) provides an excellent overview of the historical evolution of the SVD and the main contributors to its development. The SVD was proposed independently by the Italian mathematician **Eugenio Beltrami (1835–1900)** and the French mathematician **Camille Jordan (1838–1922)** in the works by Beltrami (1873) and Jordan (1874a,b). According to Stewart (1993), they were both interested in bilinear forms of the form $f(x, y) = x^\mathsf{T} Ay$, where A is a square real matrix of size $N \times N$ and x and y are column vectors. They were motivated to introduce and compute a decomposition for A in the form $A = U\Sigma V^\mathsf{T}$ in order to reduce the bilinear form to the canonical form

$$f(x,y) = \underbrace{x^\mathsf{T} U}_{\triangleq\, a^\mathsf{T}} \Sigma \underbrace{V^\mathsf{T} y}_{\triangleq\, b} = a^\mathsf{T}\Sigma b = \sum_{n=1}^{N} \sigma_n\, a_n b_n \tag{1.221}$$

in terms of the entries of $\{a, b\}$ and the diagonal entries of Σ. Beltrami (1873) focused on square and invertible matrices A while Jordan (1874a,b) considered degenerate situations with singularities as well. Beltrami (1873) exploits property (1.240) and uses it to relate the SVD factors U and V to the eigen-decompositions of the matrix products $A^\mathsf{T} A$ and AA^T. Unaware of the works by Beltrami and Jordan, the English mathematician **James Sylvester (1814–1897)** also introduced the singular value decomposition over a decade later in the works by Sylvester (1889a,b). The SVD was later extended to complex-valued matrices by Autonne (1913) and to rectangular matrices by Eckart and Young (1939). In the process, Eckart and Young (1936) re-discovered a *low-rank approximation theorem* established earlier by the same German mathematician **Erhard Schmidt (1876–1959)** of Gram-Schmidt fame in the work by Schmidt (1907), and which is nowadays known as the Eckart-Young theorem — see the account by Stewart (1993). The proof of the following statement is left to Prob. 1.56 — see also Van Huffel and Vandewalle (1987).

Eckart-Young theorem (Schmidt (1907), Eckart and Young (1936)): *Given an $N \times N$ real matrix A, consider the problem of seeking a low-rank approximation for A of rank no larger than $r < N$ by solving the problem:*

$$\widehat{A} \triangleq \underset{\{x_m, y_m\}}{\text{argmin}} \left\{ \left\| A - \sum_{m=1}^{r} x_m y_m^\mathsf{T} \right\|_\mathrm{F}^2 \right\} \tag{1.222}$$

in terms of the Frobenius norm of the difference between A and its approximation, and where $\{x_m, y_m\}$ are column vectors to be determined. If we introduce the SVD of A:

$$A = \sum_{n=1}^{N} \sigma_n u_n v_n^\mathsf{T} \tag{1.223}$$

and order the singular values $\{\sigma_n\}$ in decreasing order, i.e., $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_N$, then the solution to (1.222) is given by

$$\widehat{A} = \sum_{n=1}^{r} \sigma_n u_n v_n^\mathsf{T} \tag{1.224}$$

in terms of the singular vectors $\{u_n, v_n\}$ associated with the r largest singular values.

Today, the SVD is a widely adopted tool in scientific computing. One of the most widely used procedures for its evaluation is the algorithm proposed by Golub and Kahan (1965) and refined by Golub and Reinsch (1970). For additional discussion on the SVD and its properties, the reader may consult Horn and Johsnon (1990) and Golub and Van Loan (1996). In Appendix 1.B we provide one constructive proof for the SVD motivated by arguments from Horn and Johnson (1990), Strang (2009), Calafiore and El Ghaoui (2014), Lay, Lay, and McDonald (2014), and Nicholson (2019).

Matrix norms. A useful reference for the induced matrix norm (1.170) and, more generally, for vector and matrix norms and their properties, is Horn and Johnson (1990) — see also Golub and Van Loan (1996) and Meyer (2001). References for the block maximum norm (1.182) and (1.183) are Bertsekas and Tsitsiklis (1997), Takahashi and Yamada (2008), Takahashi, Yamada, and Sayed (2010), and Sayed (2014c); the latter reference provides several additional properties in its Appendix D and shows how this norm is useful in the study of multi-agent systems where block vector structures arise naturally.

Rayleigh-Ritz ratio. We described in Sec. 1.1 the Rayleigh-Ritz characterization of the eigenvalues of symmetric matrices, which we already know are real-valued. For an $N \times N$ real symmetric matrix, the quantity $x^{\mathsf{T}}Ax/x^{\mathsf{T}}x$ is called the Rayleigh-Ritz ratio after Ritz (1908,1909) and the Nobel Laureate in Physics **Lord Rayleigh (1842–1919)** in the works by Rayleigh (1877,1878). Both authors developed methods for determining the natural frequencies of vibrating systems (such as strings or bars). They transformed the problem of finding the natural frequencies into equivalent problems involving the determination of the stationary points of the ratio of quadratic terms. It appears that the solution method by Ritz (1908,1909) was more complete with performance guarantees and is more widely adopted. Nevertheless, both authors relied on the use of the ratio of quadratic terms, which justifies the designation Rayleigh-Ritz quotient or ratio. Accounts on the contributions of Rayleigh and Ritz are given by Courant (1943) and Leissa (2005), and by Lindsay (1945) in the introduction to the 1945 Dover editions of Rayleigh (1877,1878).

Eigenvalue perturbations. We described in Sec. 1.11 two useful results that provide bounds on the eigenvalues of matrices. The first result is Weyl Theorem, which shows how the eigenvalues of a symmetric matrix are disturbed through additive perturbations to the entries of the matrix. The second result is Gershgorin Theorem (also known as Gershgorin circle theorem or Gershogorin disc theorem), which specifies circular regions within which the eigenvalues of a matrix are located. The original references for these results are Weyl (1909,1912) and Gerschgorin (1931). A useful overview of Weyl inequality and its ramifications, along with historical remarks, appear in Bhatia (2001) and Stewart (1993); the latter reference discusses the significance of Weyl (1912) in the development of the theory of the singular value decomposition. A second useful overview of eigenvalue problems and perturbation results from the 20th Century appears in Golub and van der Vost (2001). For extensions and generalized treatments of both theorems, the reader my refer to Feingold and Varga (1962), Wilkinson (1965), Horn and Johnson (1990), Stewart and Sun (1990), Brualdi and Mellendorf (1994), Golub and Van Loan (1996), Demmel (1997), Parlett (1998), and Varga (2004).

Stochastic matrices. These matrices are prevalent in the study of Markov chains involving a finite number of states and in the study of distributed learning over graphs (see future Chapters 25 and 38). The matrices are used to represent the transition probabilities from one state to another in the Markov chain:

$$[A]_{nm} \;=\; \mathbb{P}\,(\text{transitioning from state } n \text{ to state } m) \tag{1.225}$$

Discussions on properties of stochastic matrices can be found in Minc (1988), Horn and Johnson (1990), Berman and Plemmons (1994), Meyer (2001), Seneta (2007), and Sayed (2014c, App. C). The existence of the Perron vector defined by (1.197) is guaranteed by a famous result known as the *Perron-Frobenius theorem* due to Perron (1907) and Frobenius (1908,1909,1912). The theorem applies more generally to matrices with nonnegative entries (i.e., the columns or rows of A do not need to add up to one). A useful survey appears in Pillai, Suel, and Cha (2005). To state the theorem, we first introduce the notions of irreducible and primitive matrices.

Let A denote an $N \times N$ matrix with nonnegative entries. We view each entry $a_{\ell k}$ as a weight from state ℓ to state k. The matrix A is said to be *irreducible* if, and only if, for every pair of indices (ℓ, k), there exists a finite integer $n_{\ell k} > 0$ such that

$$[A^{n_{\ell k}}]_{\ell k} > 0 \tag{1.226}$$

From the rules of matrix multiplication, the $(\ell, k)-$th entry of the $n_{\ell k}-$th power of A is given by:

$$[A^{n_{\ell k}}]_{\ell k} \;=\; \sum_{m_1=1}^{N} \sum_{m_2=1}^{N} \cdots \sum_{m_{n_{\ell k}-1}=1}^{N} a_{\ell m_1} a_{m_1 m_2} \ldots a_{m_{n_{\ell k}-1} k} \tag{1.227}$$

Therefore, property (1.226) means that there should exist a sequence of integer indexes, denoted by $(\ell, m_1, m_2, \ldots, m_{n_{\ell k}-1}, k)$, which forms a path from state ℓ to state k with $n_{\ell k}$ nonzero weights denoted by $\{a_{\ell m_1}, a_{m_1,m_2}, \ldots, a_{m_{n_{\ell k}-1},k}\}$ such that

$$\ell \overset{a_{\ell m_1}}{\longrightarrow} m_1 \overset{a_{m_1,m_2}}{\longrightarrow} m_2 \longrightarrow \ldots \longrightarrow m_{n_{\ell k}-1} \overset{a_{m_{n_{\ell k}-1},k}}{\longrightarrow} k \qquad [n_{\ell k} \text{ edges}] \tag{1.228}$$

We assume that $n_{\ell k}$ is the smallest integer that satisfies this property. Note that under irreducibility, the power $n_{\ell k}$ is allowed to be dependent on the indexes (ℓ, k). Therefore, irreducibility ensures that we can always find a path with nonzero weights linking state ℓ to state k.

A primitive matrix A is an irreducible matrix where, in addition, at least one a_{k_o,k_o} is positive for some state k_o. That is, there exists at least one state with a self-loop. It can be shown that when this holds, then an integer $n_o > 0$ exists such that — see Meyer (2001), Seneta (2007), Sayed (2014a), Prob. 1.50, and future Appendix 25.A:

$$[A^{n_o}]_{\ell k} > 0, \quad \text{uniformly for all } (\ell, k) \tag{1.229}$$

Observe that the value of n_o is now *independent* of (ℓ, k). The following statement lists some of the properties that are guaranteed by the Perron-Frobenius theorem.

Perron-Frobenius theorem (Perron (1907) and Frobenius (1908,1909,1912)): *Let A denote a square irreducible matrix with nonnegative entries and spectral radius denoted by $\lambda = \rho(A)$. Then, the following properties hold:*

(1) *$\lambda > 0$ and λ is a simple eigenvalue of A (i.e., it has multiplicity one).*
(2) *There exists a right-eigenvector, p, with all its entries positive, such that $Ap = \lambda p$.*
(3) *There exists a left-eigenvector, q, with all its entries positive, such that $A^{\mathsf{T}} q = \lambda q$.*
(4) *All other eigenvectors, associated with the other eigenvalues of A, do not share the properties of p and q, i.e., their entries are not all positive (they can have negative and/or complex entries).*
(5) *The number of eigenvalues of A whose absolute values match $\rho(A)$ is called the period of A; we denote it by the letter P. Then, all eigenvalues of A whose absolute value match $\rho(A)$ are of the form $e^{\frac{j2\pi k}{P}}\lambda$, for $k = 0, 1, \ldots, P-1$. The period P is also equal to the greatest common divisor of all integers n for which $[A^n]_{kk} > 0$.*
(6) *When A is primitive, there exists a single eigenvalue of A that matches $\rho(A)$.*

PROBLEMS

1.1 Consider a matrix $U \in \mathbb{R}^{N\times N}$ satisfying $UU^{\mathsf{T}} = I_N$. Show that $U^{\mathsf{T}}U = I_N$.
1.2 Consider a square matrix $A \in \mathbb{R}^{N\times N}$. As explained prior to (1.168), the matrix A admits the canonical Jordan decomposition $A = UJU^{-1}$, where $J = \text{blkdiag}\{J_1,\ldots,J_R\}$ is a block diagonal matrix, say, with R blocks. Each J_r has dimensions $N_r \times N_r$ where N_r represents the multiplicity of the eigenvalue λ_r. Show that $\det A = \prod_{r=1}^{R} (\lambda_r)^{N_r}$.
1.3 The trace of a square matrix is the sum of its diagonal entries. Use the canonical Jordan factorization of Prob. 1.2, and the property $\text{Tr}(XY) = \text{Tr}(YX)$ for any matrices $\{X, Y\}$ of compatible dimensions, to show that the trace of a matrix is also equal to the sum of its eigenvalues.
1.4 What are the eigenvalues of the 2×2 matrix:

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

for any angle $\theta \in [0, 2\pi]$? What are the eigenvectors of A for any $\theta \neq 0$?
1.5 Consider an arbitrary matrix $A \in \mathbb{R}^{N\times M}$. Show that its row rank is equal to its column rank. That is, show that the number of independent columns is equal to the number of independent rows (for any N and M).
1.6 Consider two square matrices $A, B \in \mathbb{R}^{N\times N}$. The matrices are said to be similar if there exists a nonsingular matrix T such that $A = TBT^{-1}$. Show that similarity transformations preserve eigenvalues, i.e., the eigenvalues of A and B coincide.
1.7 Consider an arbitrary matrix $A \in \mathbb{R}^{N\times M}$. Show that the nonzero eigenvalues of AA^{T} and $A^{\mathsf{T}}A$ coincide with each other.
1.8 Consider two symmetric and nonnegative-definite matrices A and B. Verify that $\lambda_{\min}(B)\text{Tr}(A) \leq \text{Tr}(AB) \leq \lambda_{\max}(B)\text{Tr}(A)$.
1.9 Consider two $N \times N$ matrices A and B with singular values $\{\sigma_{A,n}, \sigma_{B,n}\}$ ordered such that $\sigma_{A,1} \geq \sigma_{A,2} \geq \ldots \geq \sigma_{A,N}$ and $\sigma_{B,1} \geq \sigma_{B,2} \geq \ldots \geq \sigma_{B,N}$. Establish the following trace inequality due to von Neumann (1937): $|\text{Tr}(AB)| \leq \sum_{n=1}^{N} \sigma_{A,n}\sigma_{B,n}$.
1.10 Consider arbitrary column vectors $x, y \in \mathbb{R}^N$. Verify that

$$(I_N + xy^{\mathsf{T}})^{-1} \;=\; I_N - \frac{xy^{\mathsf{T}}}{1 + y^{\mathsf{T}}x}$$

1.11 Consider two $M\times M$ invertible matrices $\{R_a, R_b\}$ and two $M\times 1$ vectors $\{x_a, x_b\}$. In many inference problems, we will be faced with constructing a new matrix R_c and a new vector x_c using the following two relations (also known as fusion equations)

$$R_c^{-1} \;\triangleq\; R_a^{-1} + R_b^{-1}, \quad R_c^{-1}x_c \;\triangleq\; R_a^{-1}x_a + R_b^{-1}x_b$$

Show that these equations can be rewritten in any of the following equivalent forms:

$$\begin{array}{ll} R_c = R_a - R_a(R_a + R_b)^{-1}R_a, & x_c = x_a + R_cR_b^{-1}(x_b - x_a) \\ R_c = R_b - R_b(R_a + R_b)^{-1}R_b, & x_c = x_b + R_cR_a^{-1}(x_a - x_b) \end{array}$$

1.12 Verify that the vectors $\{q_m\}$ that result from the Gram-Schmidt construction (1.99a)–(1.99b) are orthonormal.
1.13 Consider an arbitrary matrix $A \in \mathbb{R}^{N\times M}$ of rank r and refer to its SVD representation (1.108a) or (1.108b). Introduce the partitioning $U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}$ and $V = \begin{bmatrix} V_1 & V_2 \end{bmatrix}$ where U_1 is $N \times r$ and V_1 is $M \times r$. Show that the matrices $\{U_1, U_2, V_1, V_2\}$ provide orthonormal basis for the four fundamental spaces $\{\mathcal{R}(A), \mathcal{N}(A^{\mathsf{T}}), \mathcal{R}(A^{\mathsf{T}}), \mathcal{N}(A)\}$.

1.14 Assume A, B are symmetric positive-definite. Show that

$$\lambda_{\max}(B^{-1}A) \;=\; \max_{x\neq 0}\left\{\frac{x^{\mathsf T}Ax}{x^{\mathsf T}Bx}\right\}$$

and that the maximum is attained when x is an eigenvector of $B^{-1}A$ that is associated with its largest eigenvalue.

1.15 Consider matrices A, B, C, and D of compatible dimensions. Show that

$$\text{Tr}(A^{\mathsf T}BCD^{\mathsf T}) \;=\; (\text{vec}(A))^{\mathsf T}(D\otimes B)\text{vec}(C)$$

1.16 Establish the singular value property (7) for Kronecker products from Table 1.1.

1.17 Establish the validity of property (10) from Table 1.1. Show that it can also be written in the equivalent form $\text{Tr}(AB) = (\text{vec}(B^*))^*\,\text{vec}(A)$ in terms of the complex conjugation operation (instead of matrix transposition).

1.18 Verify that when $B \in \mathbb{C}^{N\times N}$ is Hermitian (i.e., $B = B^*$), it holds that $\text{Tr}(AB) = (\text{vec}(B))^*\,\text{vec}(A)$.

1.19 Consider a Lyapunov recursion of the form $Z_{i+1} = AZ_iA^{\mathsf T} + B$, for $i \geq 0$ and where $Z_i \in \mathbb{R}^{N\times N}$ with square matrices $\{A, B\}$. Shown that when A is stable (i.e., when all its eigenvalues lie strictly inside the unit disc), the matrix Z_i will converge to the unique solution of the Lyapunov equation $Z = AZA^{\mathsf T} + B$.

1.20 Refer to the exponential function series (1.139). Show that $\frac{d}{dt}e^{At} = Ae^{At}$.

1.21 Show that the infinity and $p-$norms of a vector $x \in \mathbb{R}^M$ are related via $\|x\|_\infty = \lim_{p\to\infty}\|x\|_p$.

1.22 Partition a vector into sub-vectors as $x = \text{blkcol}\{x_1, x_2, \ldots, x_K\}$ and define $\|x\|_{1,p} = \sum_{k=1}^{K}\|x_k\|_p$, which is equal to the sum of the ℓ_p-norms of the sub-vectors. Show that $\|x\|_{1,p}$ is a valid vector norm.

1.23 Establish properties (a)–(e) for the block maximum norm stated right after (1.183). *Remark*. More discussion on the properties of this norm appears in Appendix D of Sayed (2014c).

1.24 Let $\|\cdot\|$ denote some norm in $\mathbb{R}^M$. The associated dual norm is denoted by $\|x\|_\star$ and defined by (1.157). Show that $\|x\|_\star$ is a valid vector norm. Show that it can be expressed equivalently by (1.158).

1.25 Let $p, q \geq 1$ satisfy $1/p + 1/q = 1$. Show that the norms $\|x\|_q$ and $\|x\|_p$ are dual of each other.

1.26 Show that:

(a) The dual of the ℓ_1-norm is the $\ell_\infty-$norm.

(b) The dual of the $\ell_\infty-$norm is the ℓ_1-norm.

1.27 Show that the dual norm of a dual norm is the original norm, i.e., $\|x\|_{\star\star} = \|x\|$.

1.28 Show that the dual norm of the Frobenius norm is the Frobenius norm itself.

1.29 Show that the dual norm of the $2-$induced norm of a matrix is the nuclear norm, which is defined as the sum of its singular values.

1.30 Refer to the matrix norms in Table 1.4. Does it hold that $\|A\| = \|A^{\mathsf T}\|$?

1.31 Show that, for any matrix norm, $|\text{Tr}(A)| \leq c\,\|A\|$ for some constant $c \geq 0$.

1.32 Let R_u denote an $M \times M$ symmetric positive-definite matrix with eigenvalues $\{\lambda_m\}$. Show that $\text{Tr}(R_u)\text{Tr}(R_u^{-1}) \geq M^2$ and $(\text{Tr}(R_u))^2 \leq M\text{Tr}(R_u^2)$.

1.33 Consider two symmetric non-negative definite matrices, $A \geq 0$ and $B \geq 0$. Show that $\text{Tr}(AB) \leq \text{Tr}(A)\,\text{Tr}(B)$.

1.34 For any symmetric matrices A and B satisfying $A \geq B \geq 0$, show that $\det A \geq \det B$. Here, the notation $A \geq B$ means that the difference $A - B$ is nonnegative definite.

1.35 Consider a symmetric $N \times N$ non-negative definite matrix, $A \geq 0$, and a second arbitrary matrix B also of size $N \times N$. Show that $|\text{Tr}(AB)| \leq \text{Tr}(A)\,\|B\|$, in terms of the $2-$induced norm (maximum singular value) of B.

1.36 Show that the spectral radius of a square symmetric matrix A agrees with its spectral norm (maximum singular value), i.e., $\rho(A) = \|A\|$.

1.37 For any matrix norm, show that the spectral radius of a square matrix A satisfies $\rho(A) \leq \|A\|$.
1.38 For any matrix norm and $\epsilon > 0$, show that $\|A^n\|^{1/n} \leq \rho(A) + \epsilon$ for n large enough.
1.39 For any matrix norm, show that the spectral radius of a square matrix A satisfies $\rho(A) = \lim_{n\to\infty} \|A^n\|^{1/n}$.
1.40 Let H denote a positive-definite symmetric matrix and let G denote a symmetric matrix of compatible dimensions. Show that $HG \geq 0$ if, and only if, $G \geq 0$, where the notation $A \geq 0$ means that all eigenvalues of matrix A are nonnegative.
1.41 Introduce the notation $\|x\|^2_\Sigma = x^\mathsf{T}\Sigma x$, where Σ is a symmetric and non-negative definite matrix.
(a) Show that $\|x\|_\Sigma = \sqrt{x^\mathsf{T}\Sigma x}$ is a valid vector norm when Σ is positive-definite, i.e., verify that it satisfies all the properties of vector norms.
(b) When Σ is singular, which properties of vector norms are violated?
1.42 Show that the condition numbers of H and $H^\mathsf{T}H$ satisfy $\kappa(H^\mathsf{T}H) = \kappa^2(H)$.
1.43 Consider a linear system of equations of the form $Ax = b$ and let $\kappa(A)$ denote the condition number of A. Assume A is invertible so that the solution is given by $x = A^{-1}b$. Now assume that b is perturbed slightly to $b + \delta b$. The new solution becomes $x + \delta x$, where $\delta x = A^{-1}\delta b$. The relative change in b is $\beta = \|\delta b\|/\|b\|$. The relative change in the solution is $\alpha = \|\delta x\|/\|x\|$. When the matrix is ill-conditioned, the relative change in the solution can be much larger than the relative change in b. Indeed, verify that $\alpha/\beta \leq \kappa(A)$. Can you provide an example where the ratio achieves $\kappa(A)$?
1.44 Consider a matrix Y of the form $Y = I - \beta xx^\mathsf{T}$, for some real scalar β. For what condition on β is Y positive semi-definite? When this is the case, show that Y admits a symmetric square-root factor of the form $Y^{1/2} = I - \alpha xx^\mathsf{T}$ for some real scalar α.
1.45 Let $\Phi = \sum_{n=0}^{N} \lambda^{N-n} h_n h_n^\mathsf{T}$, where the h_n are $M \times 1$ vectors and $0 \ll \lambda \leq 1$. The matrix Φ can have full-rank or it may be rank deficient. Assume its rank is $r \leq M$. Let $\Phi^{1/2}$ denote an $M \times r$ full-rank square-root factor, i.e., $\Phi^{1/2}$ has rank r and satisfies $\Phi^{1/2}\Phi^{\mathsf{T}/2} = \Phi$. Show that h_n belongs to the column span of $\Phi^{1/2}$.
1.46 Refer to Example 1.9. Extend the result to the case $A^\mathsf{T}A = B^\mathsf{T}B$ where now $N \geq M$.
1.47 We provide another derivation for the basis rotation result from Example 1.9 by assuming that the matrices A and B have full rank. Introduce the QR decompositions

$$A^\mathsf{T} = Q_A \begin{bmatrix} R_A \\ 0 \end{bmatrix}, \qquad B^\mathsf{T} = Q_B \begin{bmatrix} R_B \\ 0 \end{bmatrix}$$

where Q_A and Q_B are $M \times M$ orthogonal matrices, and R_A and R_B are $N \times N$ upper triangular matrices with positive diagonal entries (due to the full rank assumption on A and B).
(a) Show that $AA^\mathsf{T} = R_A^\mathsf{T} R_A = R_B^\mathsf{T} R_B$.
(b) Conclude, by uniqueness of the Cholesky factorization, that $R_A = R_B$. Verify further that $\Theta = Q_B Q_A^\mathsf{T}$ is orthogonal and maps B to A.
1.48 Consider a matrix $B \in \mathbb{R}^{N\times M}$ and let $\sigma^2_{\min}(B)$ denote its smallest nonzero singular value. Let $x \in \mathcal{R}(B)$ (i.e., x is any vector in the range space of B). Use the eigenvalue decomposition of $B^\mathsf{T}B$ to verify that $\|B^\mathsf{T}x\|^2 \geq \sigma^2_{\min}(B)\,\|x\|^2$.
1.49 Refer to Sec. 1.12 on stochastic matrices. Show that the spectral radius of a left or right-stochastic matrix is equal to one.
1.50 Let A denote an irreducible matrix with nonnegative entries. Show that if $a_{k_o,k_o} > 0$ for some k_o, then A is a primitive matrix.
1.51 Let A denote a matrix with positive entries. Show that A is primitive.
1.52 Show that to check whether an $N \times N$ left-stochastic matrix A is irreducible (primitive), we can replace all nonzero entries in A by ones and verify instead whether the resulting matrix is irreducible (primitive).
1.53 Consider an $N \times N$ left-stochastic matrix A that is irreducible but not necessarily

primitive. Let $B = 0.5(I + A)$. Is B left-stochastic? Show that the entries of B^{N-1} are all positive. Conclude that B is primitive.

1.54 Assume A is a left-stochastic primitive matrix of size $N \times N$.

(a) Show that A is power convergent and the limit converges to the rank-one product $\lim_{n\to\infty} A^n = p\mathbb{1}^{\mathsf{T}}$, where p is the Perron vector of A. Is the limit a primitive matrix?

(b) For any vector $b = \text{col}\{b_1, b_2, \ldots, b_N\}$, show that $\lim_{n\to\infty} A^n b = \alpha p$, where $\alpha = b_1 + b_2 + \ldots + b_N$.

(c) If A is irreducible but not necessarily primitive, does the limit of part (a) exist?

1.55 Consider an $N \times N$ left-stochastic matrix A. Let $n_o = N^2 - 2N + 2$. Show that A is primitive if, and only if, $[A^{n_o}]_{\ell k} > 0$ for all ℓ and k.

1.56 Establish the validity of the Eckart-Young approximation expression (1.224).

1.57 Consider an $M \times N$ matrix A (usually $N \geq M$ with more columns than rows). The *spark* of A is defined as the smallest number of linearly *dependent* columns in A, also written as:

$$\text{spark}(A) \;\stackrel{\Delta}{=}\; \min_{d\neq 0} \Big\{\|d\|_0\Big\}, \quad \text{subject to } Ad = 0$$

where $\|d\|_0$ denotes the number of nonzero arguments in the vector d. If A has full rank, then its spark is set to ∞. Now consider the linear system of equations $Ax^o = b$ and assume x^o is $k-$sparse in the sense that only k of its entries are nonzero. Show that x^o is the only solution to the following problem (i.e., the only $k-$sparse vector satisfying the linear equations):

$$x^o = \underset{x\in\mathbb{R}^M}{\text{argmin}} \Big\{\|x\|_0\Big\}, \;\; \text{subject to } Ax = b$$

if, and only if, $\text{spark}(A) > 2k$. *Remark.* We describe later in Appendix 58.A the *orthogonal matching pursuit* (OMP) algorithm for finding the sparse solution x^o.

1.58 Consider an $M \times N$ matrix A. The matrix is said to satisfy a *restricted isometry property* (RIP) with constant λ_k if for any $k-$sparse vector x it holds $(1-\lambda_k)\|x\|_2 \leq \|Ax\|_2 \leq (1+\lambda_k)\|x\|_2$. Now let $Ax_1 = b_1$ and $Ax_2 = b_2$. Assume A satisfies RIP for any sparse vectors of level $2k$. Show that for any $k-$sparse vectors $\{x_1, x_2\}$, the corresponding measurements $\{b_1, b_2\}$ will be sufficiently away from each other in the sense that $\|b_1 - b_2\|_2 \geq (1 - \lambda_{2k})\|x_1 - x_2\|_2$. *Remark.* For more discussion on sparsity and the RIP condition, the reader may refer to the text by Elad (2010) and the many references therein, as well as to the works by Candes and Tao (2006) and Candes, Romberg, and Tao (2006).

1.A PROOF OF SPECTRAL THEOREM

The arguments in this appendix and the next on the spectral theorem and the singular value decomposition are motivated by the presentations in Horn and Johnson (1990), Trefethen and Bau (1997), Strang (2009), Calafiore and El Ghaoui (2014), Lay, Lay, and McDonald (2014), and Nicholson (2019).

In this first appendix, we establish the validity of the eigen-decomposition (1.15a) for $N \times N$ real symmetric matrices A. We start by verifying that A has at least one real eigenvector. For this purpose, we first recall that an equivalent characterization of the eigenvalues of a matrix is that they are the roots of its characteristic polynomial, defined as

$$p(\lambda) \;\stackrel{\Delta}{=}\; \det(\lambda I_N - A) \tag{1.230}$$

in terms of the determinant of the matrix $\lambda I_N - A$. Note that $p(\lambda)$ is a polynomial of order N and, by the fundamental theorem of algebra, every such polynomial has N

roots. We already know that these roots are all real when A is symmetric. Therefore, there exists some real value λ_1 such that $p(\lambda_1) = 0$. The scalar λ_1 makes the matrix $\lambda_1 I_N - A$ singular since its determinant will be zero. In this way, the columns of the matrix $\lambda_1 I_N - A$ will be linearly dependent and there must exist some nonzero real vector u_1 such that

$$(\lambda_1 I_N - A)u_1 = 0 \tag{1.231}$$

This relation establishes the claim that there exists some nonzero real vector u_1 satisfying $Au_1 = \lambda_1 u_1$. We can always scale u_1 to satisfy $\|u_1\| = 1$.

Induction argument

One traditional approach to establish the spectral theorem is by induction. Assume the theorem holds for all symmetric matrices of dimensions up to $(N-1)\times(N-1)$ and let us prove that the statement also holds for the next dimension $N \times N$; it certainly holds when $N = 1$ (in which case A is a scalar). Thus, given an $N \times N$ real symmetric matrix A, we already know that it has at least one real eigenvector u_1 of unit-norm associated with a real eigenvalue λ_1. Since u_1 lies in $N-$dimensional space, we can choose $N-1$ real vectors $\{\bar{v}_2, \bar{v}_3, \ldots, \bar{v}_N\}$ such that the columns of the $N \times N$ matrix

$$\bar{V} = \left[\begin{array}{c|cccc} u_1 & \bar{v}_2 & \bar{v}_3 & \ldots & \bar{v}_N \end{array}\right] \tag{1.232}$$

are linearly independent. The columns of $\bar{V}$ constitute a basis for the $N-$dimensional space, $\mathbb{R}^N$. We apply the Gram-Schmidt orthogonalization procedure to the trailing columns of $\bar{V}$ and replace the $\{\bar{v}_n\}$ by a new set of vectors $\{v_n\}$ that have unit norm each, and such that the columns of the matrix V below are all orthogonal to each other:

$$V = \left[\begin{array}{c|cccc} u_1 & v_2 & v_3 & \ldots & v_N \end{array}\right] \overset{\Delta}{=} \left[\begin{array}{cc} u_1 & V_1 \end{array}\right] \tag{1.233}$$

Note that we kept u_1 unchanged; we are also denoting the trailing columns of V by V_1. Now, multiplying A by $V^{\mathsf T}$ from the left and by V from the right we get

$$V^{\mathsf T}AV = \begin{bmatrix} u_1^{\mathsf T}Au_1 & u_1^{\mathsf T}AV_1 \\ V_1^{\mathsf T}Au_1 & V_1^{\mathsf T}AV_1 \end{bmatrix} = \begin{bmatrix} \lambda_1\|u_1\|^2 & \lambda_1 u_1^{\mathsf T}V_1 \\ \lambda_1 V_1^{\mathsf T}u_1 & V_1^{\mathsf T}AV_1 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & V_1^{\mathsf T}AV_1 \end{bmatrix} \tag{1.234}$$

since $\|u_1\|^2 = 1$ and u_1 is orthogonal to the columns of V_1 (so that $V_1^{\mathsf T}u_1 = 0$). Note that we used the fact that A is symmetric in the above calculation to conclude that

$$u_1^{\mathsf T}A \;=\; u_1^{\mathsf T}A^{\mathsf T} \;=\; (Au_1)^{\mathsf T} \;=\; (\lambda_1 u_1)^{\mathsf T} \;=\; \lambda_1 u_1^{\mathsf T} \tag{1.235}$$

Thus, observe that the matrix product $V^{\mathsf T}AV$ turns out to be block diagonal with λ_1 as its $(1,1)$ leading entry and with the $(N-1)\times(N-1)$ real symmetric matrix, $V_1^{\mathsf T}AV_1$, as its trailing block. We know from the induction assumption that this smaller symmetric matrix admits a full set of orthonormal eigenvectors. That is, there exists an $(N-1)\times(N-1)$ real orthogonal matrix W_1 and a diagonal matrix Λ_1 with real entries such that

$$V_1^{\mathsf T}AV_1 \;=\; W_1\Lambda_1 W_1^{\mathsf T}, \quad W_1^{\mathsf T}W_1 = I_{N-1} \tag{1.236}$$

or, equivalently,

$$W_1^{\mathsf T}V_1^{\mathsf T}AV_1W_1 \;=\; \Lambda_1 \tag{1.237}$$

Using this equality, we obtain from (1.234) that

$$\underbrace{\begin{bmatrix} 1 & \\ & W_1^{\mathsf T} \end{bmatrix} V^{\mathsf T}}_{=\,U^{\mathsf T}} A \underbrace{V \begin{bmatrix} 1 & \\ & W_1 \end{bmatrix}}_{\overset{\Delta}{=}\,U} = \underbrace{\begin{bmatrix} \lambda_1 & \\ & \Lambda_1 \end{bmatrix}}_{\overset{\Delta}{=}\,\Lambda} \tag{1.238}$$

with a diagonal matrix on the right-hand side. The matrix U can be verified to be orthogonal since

$$U^{\mathsf{T}}U = \begin{bmatrix} 1 & \\ & W_1^{\mathsf{T}} \end{bmatrix} \underbrace{V^{\mathsf{T}}V}_{=I} \begin{bmatrix} 1 & \\ & W_1 \end{bmatrix} = \begin{bmatrix} 1 & \\ & W_1^{\mathsf{T}}W_1 \end{bmatrix} = I_N \tag{1.239}$$

Therefore, we established that an orthogonal matrix U and a real diagonal matrix Λ exist such that $A = U\Lambda U^{\mathsf{T}}$, as claimed.

1.B CONSTRUCTIVE PROOF OF SVD

One proof of the singular value decomposition defined by (1.108a)–(1.108b) follows from the eigen-decomposition of symmetric nonnegative-definite matrices. The argument given here assumes $N \leq M$, but it can be adjusted to handle the case $N \geq M$. Thus, note that the product AA^{T} is a symmetric nonnegative-definite matrix of size $N \times N$. Consequently, from the spectral theorem, there exists an $N \times N$ orthogonal matrix U and an $N \times N$ diagonal matrix Σ^2, with nonnegative entries, such that

$$AA^{\mathsf{T}} = U\Sigma^2 U^{\mathsf{T}} \tag{1.240}$$

This representation corresponds to the eigen-decomposition of AA^{T}. The diagonal entries of Σ^2 are the eigenvalues of AA^{T}, which are nonnegative (and, hence, the notation Σ^2); the nonzero entries of Σ^2 also coincide with the nonzero eigenvalues of $A^{\mathsf{T}}A$ — see Prob. 1.7. The columns of U are the orthonormal eigenvectors of AA^{T}. By proper reordering, we can arrange the diagonal entries of Σ^2, denoted by $\{\sigma_r^2\}$, in decreasing order so that Σ^2 can be put into the form

$$\Sigma^2 = \text{diagonal}\left\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_r^2, 0, \ldots, 0\right\} \triangleq \begin{bmatrix} \Lambda^2 & \\ & 0_{N-r} \end{bmatrix} \tag{1.241}$$

where $r = \text{rank}(AA^{\mathsf{T}})$ and $\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_r^2 > 0$. The $r \times r$ diagonal matrix Λ consists of the positive entries

$$\Lambda \triangleq \text{diagonal}\{\sigma_1, \sigma_2, \ldots, \sigma_r\} > 0 \tag{1.242}$$

We partition $U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}$, where U_1 is $N \times r$, and conclude from the orthogonality of U (i.e., from $U^{\mathsf{T}}U = I$) that $U_1^{\mathsf{T}}U_1 = I_r$ and $U_1^{\mathsf{T}}U_2 = 0$. If we substitute into (1.240) we find that

$$AA^{\mathsf{T}} = U_1\Lambda^2 U_1^{\mathsf{T}} \tag{1.243a}$$

$$AA^{\mathsf{T}}U = U\Sigma^2 \implies AA^{\mathsf{T}}U_2 = 0 \overset{(1.50)}{\iff} A^{\mathsf{T}}U_2 = 0 \tag{1.243b}$$

where the middle expression in the last line is indicating that the columns of U_2 belong to the nullspace of AA^{T}. But since we know from property (1.50) that $\mathcal{N}(AA^{\mathsf{T}}) = \mathcal{N}(A^{\mathsf{T}})$, we conclude that the columns of U_2 also belong to the nullspace of A^{T}. Next, we introduce the $M \times r$ matrix

$$V_1 \triangleq A^{\mathsf{T}}U_1\,\Lambda^{-1} \tag{1.244}$$

The columns of V_1 are orthonormal since

$$V_1^{\mathsf{T}}V_1 = \underbrace{\Lambda^{-1}U_1^{\mathsf{T}}A}_{=V_1^{\mathsf{T}}}\underbrace{A^{\mathsf{T}}U_1\Lambda^{-1}}_{V_1} = \Lambda^{-1}U_1^{\mathsf{T}}\underbrace{U_1\Lambda^2U_1^{\mathsf{T}}}_{=AA^{\mathsf{T}}}U_1\Lambda^{-1} = I_r \tag{1.245a}$$

Moreover, it also holds that

$$V_1^\mathsf{T} A^\mathsf{T} U_1 = \underbrace{\Lambda^{-1} U_1^\mathsf{T} A}_{=V_1^\mathsf{T}} A^\mathsf{T} U_1 \;=\; \Lambda^{-1} U_1^\mathsf{T} \underbrace{U_1 \Lambda^2 U_1^\mathsf{T}}_{=AA^\mathsf{T}} U_1 \;=\; \Lambda^{-1}\Lambda^2 \;=\; \Lambda \tag{1.245b}$$

and, similarly,

$$V_1^\mathsf{T} A^\mathsf{T} U_2 = 0_{r\times(N-r)}, \quad \text{since } A^\mathsf{T} U_2 = 0 \tag{1.245c}$$

Combining (1.245b)–(1.245c) gives

$$V_1^\mathsf{T} A^\mathsf{T} U = \begin{bmatrix} \Lambda & 0_{r\times(N-r)} \end{bmatrix} \tag{1.246}$$

Now, since V_1 has r orthonormal columns in $M-$dimensional space, we can add $M-r$ more columns to enlarge V_1 into an $M \times M$ orthogonal matrix V as follows:

$$V \;\stackrel{\Delta}{=}\; \begin{bmatrix} V_1 & V_2 \end{bmatrix}, \;\; V^\mathsf{T} V = I_M \tag{1.247}$$

which implies that the new columns in $V_2 \in \mathbb{R}^{M\times(M-r)}$ should satisfy $V_2^\mathsf{T} V_1 = 0$ and $V_2^\mathsf{T} V_2 = I$. It follows that

$$\begin{aligned} V_2^\mathsf{T} V_1 = 0 &\Longrightarrow V_2^\mathsf{T} \underbrace{A^\mathsf{T} U_1 \Lambda^{-1}}_{=V_1} = 0 \\ &\Longrightarrow V_2^\mathsf{T} A^\mathsf{T} U_1 = 0_{(M-r)\times r}, \quad \text{since } \Lambda > 0 \end{aligned} \tag{1.248}$$

and

$$V_2^\mathsf{T} A^\mathsf{T} U_2 = 0_{(M-r)\times(N-r)}, \quad \text{since } A^\mathsf{T} U_2 = 0. \tag{1.249}$$

Adding these conclusions into (1.246) we can write

$$\begin{bmatrix} V_1^\mathsf{T} \\ V_2^\mathsf{T} \end{bmatrix} A^\mathsf{T} U = \begin{bmatrix} \Lambda & 0_{r\times(N-r)} \\ 0_{(M-r)\times r} & 0_{(M-r)\times(N-r)} \end{bmatrix} \tag{1.250}$$

and we conclude that orthogonal matrices U and V exist such that

$$A = U \underbrace{\begin{bmatrix} \Lambda & 0_{r\times(M-r)} \\ 0_{(N-r)\times r} & 0_{(N-r)\times(M-r)} \end{bmatrix}}_{=\begin{bmatrix} \Sigma & 0_{N\times(M-N)} \end{bmatrix}} V^\mathsf{T} \tag{1.251}$$

as claimed by (1.108a). A similar argument establishes the SVD decomposition of A when $N > M$.

REFERENCES

Aitken, A. C. (1939), *Determinants and Matrices*, 9th edition, Interscience Publishers, NY.

Autonne, L. (1913), "Sur les matrices hypohermitiennes et les unitairs," *Comptes Rendus de l'Academie Sciences*, Paris, vol. 156, pp. 858–860.

Banachiewicz, T. (1937a), "Sur l'inverse d'un cracovien et une solution générale d'un systeme d'équations linéaires," *Comptes Rendus Mensules des Séances de la Classe des Sciences Mathématique et Naturelles de l'Académie Polonaise des Sciences et des Lettres*, no. 4, pp. 3–4.

Banachiewicz, T. (1937b), "Zur Berechungung der Determinanten, wie auch der Inversen, und zur darauf basierten Auflosung der Systeme lineare Gleichungen," *Acta Astronomica*, Sér. C, vol. 3, pp. 41–67.

Bartlett, M. S. (1951), "An inverse matrix adjustment arising in discriminant analysis," *Ann. Math. Statist.,* vol. 22, pp. 107–111.

Bellman, R. E. (1970), *Introduction to Matrix Analysis*, 2nd edition, McGraw Hill, NY.

Beltrami, E. (1873), "Sulle funzioni bilineari," *Giornale di Matematiche ad Uso degli Studenti Delle Universita*, vol. 11, pp. 98–110.

Berman, A. and R. J. Plemmons (1994), *Nonnegative Matrices in the Mathematical Sciences*, SIAM, PA.

Bernstein, D. S. (2018), *Scalar, Vector, and Matrix Mathematics: Theory, Facts, and Formulas*, revised edition, Princeton University Press, NJ.

Bertsekas, D. P. and J. N. Tsitsiklis (1997), *Parallel and Distributed Computation: Numerical Methods*, Athena Scientific, Singapore.

Bhatia, R. (2001), "Linear algebra to quantum cohomology: The story of Alfred Horn's inequalities," *The American Mathematical Monthly*, vol. 108, no. 4, pp. 289–318.

Bjorck, A. (1996), *Numerical Methods for Least Squares Problems,* SIAM, PA.

Bródy, F. and T. Vámos (1995), *The Neumann Compendium*, vol. 1, World Scientific, Singapore.

Brualdi, R. A. and S. Mellendorf (1994), "Regions in the complex plane containing the eigenvalues of a matrix," *Amer. Math. Monthly*, vol. 101, pp. 975–985.

Calafiore, G. C. and L. El Ghaoui (2014), *Optimization Models*, Cambridge University Press.

Candes, E. J., J. K. Romberg, and T. Tao (2006), "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Information Theory*, vol. 52, no. 2, pp. 489–509.

Candes, E. J. and T. Tao (2006), "Near-optimal signal recovery from random projections: Universal encoding strategies," *IEEE Trans. Information Theory*, vol. 52, pp. 5406–5425.

Carrera, J. P. (1992), "The fundamental theorem of algebra before Carl Friedrich Gauss," *Publ. Mat.*, vol. 36, pp. 879–911.

Cauchy, A.-L. (1829), "Sur l'équation a l'aide de laquelle on determine les inégalités séculaires des mouvements des planetes," *Exer. de math.*, vol. 4, pp. 174–195.

Constantinescu, T. (1996), *Schur Parameters, Factorization, and Dilation Problems*, Birkhaüser, Berlin.

Cottle, R. W. (1974), "Manifestations of the Schur complement," *Linear Algebra and Its Appl.*, vol. 8, pp. 189–211.

Courant, R. (1943), "Variational methods for the solution of problems of equilibrium and vibrations," *Bull. Amer. Math. Soc.*, vol. 49, pp. 1–23.

Demmel, J. (1997), *Applied Numerical Linear Algebra*, SIAM, PA.

Duncan, W. J. (1944), "Some of the solution of large sets of simultaneous linear equations (with an appendix on the reciprocation of partitioned matrices)," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, Seventh Series, vol. 35, pp. 660–670.

Eckart, C. and G. Young (1936), "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218.

Eckart, C. and G. Young (1939), "A principal axis transformation for non-Hermitian matrices," *Bull. Amer. Math. Soc.*, vol. 45, pp. 118–121.

Elad, M. (2010), *Sparse and Redundant Representations*, Springer, NY.

Feingold, D. G. and R. S. Varga (1962), "Block diagonally dominant matrices and generalizations of the Gerschgorin circle theorem," *Pacific J. Math.*, vol. 12, pp. 1241–1250.

Fine, B. and G. Rosenberger (1997), *The Fundamental Theorem of Algebra*, Springer-Verlag, NY.

Frobenius, F. G. (1908), "Uber matrizen aus positiven elementen, 1" *Sitzungsber. Konigl. Preuss. Akad. Wiss.*, pp. 471–476.

Frobenius, G. (1909), "Uber matrizen aus positiven elementen, 2," *Sitzungsber. Konigl. Preuss. Akad. Wiss.*, pp. 514–518.

Frobenius, G. (1912), "Uber matrizen aus nicht negativen elementen," *Sitzungsber. Konigl. Preuss. Akad. Wiss.*, pp. 456–477.

Gantmacher, F. R. (1959), *The Theory of Matrices*, Chelsea Publishing Company, NY.

Gelfand, I. M. (1989), *Lectures on Linear Algebra*, Dover Publications, NY.

Gerschgorin, S. (1931), "Über die abgrenzung der eigenwerte einer matrix," *Izv. Akad. Nauk. USSR Otd. Fiz.-Mat. Nauk*, vol. 7, pp. 749–754.

Golub, G. H. and W. Kahan (1965), "Calculating the singular values and pseudo-inverse of a matrix," *J. Society Indust. Appl. Math.: Series B, Numerical Analysis*, vol. 2, no. 2, pp. 205–224.

Golub, G. H. and C. Reinsch (1970), "Singular value decomposition and least squares solutions," *Numerische Mathematik*, vol. 14, no. 5, pp. 403–420.

Golub, G. H. and H. A. van der Vost (2001), "Eigenvalue computation in the 20th century," in *Numerical Analysis: Historical Developments in the 20th Century*, C. Brezinski and L. Wuytack, *Eds.*, pp. 209–238, North-Holland, Elsevier.

Golub, G. H. and C. F. Van Loan (1996), *Matrix Computations*, 3rd edition, The John Hopkins University Press, MD.

Gram, J. (1883), "Ueber die Entwickelung reeller Funtionen in Reihen mittelst der Methode der kleinsten Quadrate," *J. Reine Angew. Math.*, vol. 94, pp. 41–73.

Guttman, L. (1946), "Enlargement methods for computing the inverse matrix," *Ann. Math. Statist.*, vol. 17, pp. 336–343.

Hackbusch, W. (2012), *Tensor Spaces and Numerical Tensor Calculus*, Springer-Verlag.

Hager, W. W. (1989), "Updating the inverse of a matrix," *SIAM Review*, vol. 31, no. 2, pp. 221–239.

Halmos, P. R. (1963), "What does the spectral theorem say?" *The American Mathematical Monthly,* vol. 70, no. 3, pp. 241–247.

Halmos, P. R. (1974), *Finite-Dimensional Vector Spaces*, Springer, NY.

Halmos, P. R. (2013), *Introduction to Hilbert Space and the Theory of Spectral Multiplicity*, Martino Fine Books, CT, USA.

Hawkins, T. (1975), "Cauchy and the spectral theory of matrices," *Historia Mathematica*, vol. 2, no. 1, pp. 1–29.

Haynsworth, E. V. (1968), "Determination of the inertia of a partitioned Hermitian matrix," *Linear Algebra and its Applications*, vol. 1, pp. 73–81.

Henderson, H. V., F. Pukelsheim, and S. R. Searle (1983), "On the history of the Kronecker product," *Linear and Multilinear Algebra*, vol. 14, pp. 113–120.

Henderson, H. V. and S. R. Searle (1981a), "On deriving the inverse of a sum of matrices," *SIAM Review*, vol. 23, pp. 53–60.

Henderson, H. V. and S. R. Searle (1981b), "The vec-permutation matrix, the vec operator, and Kronecker products: A review," *Linear and Mulitilinear Algebra*, vol. 9, pp. 271–288.

Hensel, K. (1891), "Uber die Darstellung der Determinante eines Systems, welches aus zwei anderen componirt ist," *ACTA Mathematica*, vol. 14, pp. 317–319.

Ho, Y. C. (1963), "On the stochastic approximation method and optimal filter theory," *J. Math. Anal. Appl.*, vol. 6, pp. 152–154.

Hogben, L., *Ed.* (2014), *Handbook of Linear Algebra*, 2nd edition, CRC Press.

Horn, R. A. and C. R. Johnson (1990), *Matrix Analysis*, Cambridge University Press.

Hotelling, H. (1943a), "Some new methods in matrix calculation," *Ann. Math. Statist.*, vol. 14, pp. 1–34.

Hotelling, H. (1943b), "Further points on matrix calculation and simultaneous equations," *Ann. Math. Statist.*, vol. 14, pp. 440–441.

Householder, A. S. (1953), *Principles of Numerical Analysis,* McGraw-Hill, NY.

Householder, A. S. (1957), "A survey of closed methods for inverting matrices," *J. Soc. Ind. Appl. Math.*, vol. 5, pp. 155–169.

Householder, A. S. (1964), *The Theory of Matrices in Numerical Analysis,* Blaisdell, NY.

Jordan, C. (1874a), "Mémoire sur les formes bilinéaires," *J. Math. Pures Appl.*, Deuxieme Série, vol. 19, pp. 35–54.

Jordan, C. (1874b), "Sur la réduction des formes bilinéaires," *Comptes Rendus de l'Academie Sciences*, Paris, pp. 614–617.
Kailath, T. (1960), "Estimating filters for linear time-invariant channels," *Quarterly Progress Report 58*, MIT Research Laboratory of Electronics, Cambridge, MA, pp. 185–197.
Kailath, T. (1986), "A theorem of I. Schur and its impact on modern signal processing," in *Operator Theory: Advances and Applications*, vol. 18, pp. 9–30, Birkhauser, Basel.
Kailath, T. and A. H. Sayed (1995), "Displacement structure: Theory and applications," *SIAM Review*, vol. 37, no. 3, pp. 297–386.
Koning, R. H., H. Neudecker, and T. Wansbeek (1991), "Block Kronecker products and the vecb operator," *Linear Algebra Appl.*, vol. 149, pp. 165–184.
Laplace, P. S. (1812), *Théorie Analytique des Probabilités*, Paris.
Lax, P. (1997), *Linear Algebra*, Wiley, NY.
Lay, D. (1994), *Linear Algebra and Its Applications*, Addison-Wesley, Reading, MA.
Lay, D., S. Lay, and J. McDonald (2014), *Linear Algebra and Its Applications*, 5th edition, Pearson.
Laub, A. J. (2004), *Matrix Analysis for Scientists and Engineers*, SIAM, PA.
Leissa, A. W. (2005), "The historical bases of the Rayleigh and Ritz methods," *J. Sound and Vibration*, vol. 287, pp. 961–978.
Lindsay, R. B. (1945), "Historical introduction," in J. W. S. Rayleigh, *The Theory of Sound*, vol. 1, Dover edition, NY.
Liu, S. (1999), "Matrix results on the Khatri-Rao and Tracy-Singh products," *Linear Algebra and its Applications*, vol. 289, no. 1–3, pp. 267–277.
MacDuffee, C. C. (1946), *The Theory of Matrices*, Chelsea, NY.
Meyer, C. D. (2001), *Matrix Analysis and Applied Linear Algebra*, SIAM, PA.
Minc, H. (1988), *Nonnegative Matrices*, Wiley, NY.
Nicholson, W. K. (2019), *Linear Algebra with Applications*, open edition, revision A. Available online under the Creative Commons License on lyryx.com
Parlett, B. N. (1998), *The Symmetric Eigenvalue Problem*, SIAM, PA.
Perron, O. (1907), "Zur theorie der matrices," *Mathematische Annalen*, vol. 64, no. 2, pp. 248–263.
Pillai, S. U., T. Suel, and S. Cha (2005), "The Perron-Frobenius theorem: Some of its applications," *IEEE Signal Process. Mag.*, vol. 22, no. 2, pp. 62–75.
Plackett, R. L. (1950), "Some theorems in least-squares," *Biometrika*, vol. 37, no. 1–2, pp. 149–157.
Puntanen, S. and G. P. H. Styan (2005), "Historical introduction: Issai Schur and the early development of the Schur complement," in *The Schur Complement and Its Applications*, pp. 1–16, F. Zhang, *Ed.*, Springer, NY.
Rayleigh, J. W. S. (1877), *The Theory of Sound*, vol. 1, The Macmillan Company. Reprinted in 1945 by Dover Publications, NY.
Rayleigh, J. W. S. (1878), *The Theory of Sound*, vol. 2, The Macmillan Company. Reprinted in 1945 by Dover Publications, NY.
Regalia, P. A. and S. Mitra (1989), "Kronecker products, unitary matrices, and signal processing applications," *SIAM Review*, vol. 31, pp. 586–613.
Ritz, W. (1908), "Uber eine neue methode zur Losung gewisser variationsprobleme der mathematischen," *Physik, Journal fur Reine und Angewandte Mathematik*, vol. 135, pp. 1–61.
Ritz, W. (1909), "Theorie der transversalschwingungen einer quadratische platte mit freien Randern," *Annalen der Physik*, vol. 28, pp. 737–786.
Sayed, A. H. (2003), *Fundamentals of Adaptive Filtering*, Wiley, NJ.
Sayed, A. H. (2008), *Adaptive Filters*, Wiley, NJ.
Sayed, A. H. (2014a), *Adaptation, Learning, and Optimization over Networks,* Foundations and Trends in Machine Learning, NOW Publishers, vol. 7, no. 4–5, pp. 311–801.
Sayed, A. H. (2014c), "Diffusion adaptation over networks," in *E-Reference Signal Processing*, R. Chellapa and S. Theodoridis, *Eds.*, vol. 3, pp. 323–454, Academic Press, NY.

Schmidt, E. (1907), "Zur Theorie der linearen und nichtlinearen Integralgleichungen. I. Teil. Entwicklung willkurlichen Funktionen nach System vorgeschriebener," *Math. Ann.*, vol. 63, pp. 433–476.
Schmidt, E. (1908), "Uber die Auflosung linearen Gleichungen mit unendlich vielen Unbekanten," *Rend. Circ. Mat. Palermo. Ser.* 1, vol. 25, pp. 53–77.
Schur, I. (1917), "Über potenzreihen die inm inneren des einheitskreises beschränkt sind," *Journal für die Reine und Angewandte Mathematik*, vol. 147, pp. 205–232. [English translation in *Operator Theory: Advances and Applications*, vol. 18, pp. 31–88, edited by I. Gohberg, Birkhaüser, Boston, 1986.]
Seneta, E. (2007), *Non-negative Matrices and Markov Chains*, 2nd edition, Springer, NY.
Sherman, J. and W. J. Morrison (1949), "Adjustment of an inverse matrix corresponding to changes in the elements of a given column or a given row of the original matrix," *Ann. Math. Statistics*, vol. 20, p. 621.
Sherman, J. and W. J. Morrison (1950), "Adjustment of an inverse matrix corresponding to a change in one element of a given matrix," *Ann. Math. Statist.*, vol. 21, pp. 124–127.
Stewart, G. W. (1993), "On the early history of the singular value decomposition," *SIAM Review*, vol. 35, pp. 551-566.
Stewart, G. W. and J.-G. Sun (1990), *Matrix Perturbation Theory*, Academic Press, Boston.
Strang, G. (1988), *Linear Algebra and Its Applications*, 3rd edition, Academic Press, NY.
Strang, G. (2009), *Introduction to Linear Algebra*, 4th edition, Wellesley-Cambridge Press, Wellesly, MA.
Sylvester, J. J. (1852), "A demonstration of the theorem that every homogeneous quadratic polynomial is reducible by real orthogonal substitutions to the form of a sum of positive and negative squares," *Phil. Mag.*, Ser. 4, vol. 4, no. 23, pp. 138–142.
Sylvester, J. J. (1889a), "A new proof that a general quadratic may be reduced to its canonical form (that is, a linear function of squares) by means of a real orthogonal substitution," *Messenger of Mathematics*, vol. 19, pp. 1–5.
Sylvester, J. J. (1889b), "On the reduction of a bilinear quantic of the $n-$th order to the form of a sum of n products by a double orthogonal substitution," *Messenger of Mathematics*, vol. 19, pp. 42–46.
Takahashi, N. and I. Yamada (2008), "Parallel algorithms for variational inequalities over the cartesian product of the intersections of the fixed point sets of nonexpansive mappings," *J. Approx. Theory*, vol. 153, no. 2, pp. 139–160.
Takahashi, N., I. Yamada, and A. H. Sayed (2010), "Diffusion least-mean-squares with adaptive combiners: Formulation and performance analysis," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4795–4810.
Tracy, D. S. and R. P. Singh (1972), "A new matrix product and its applications in matrix differentiation," *Statistica Neerlandica*, vol. 26, no. 4, pp. 143–157.
Trefethen, L. N. and D. Bau (1997), *Numerical Linear Algebra*, SIAM, PA.
Van Huffel, S. and J. Vandewalle (1987), *The Total Least Squares Problem: Computational Aspects and Analysis*, SIAM, PA.
Van Loan, C. F. (2000), "The ubiquitous Kronecker product," *J. Comput. Applied Math.*, vol. 123, no. 1–2, pp. 85–100.
Varga, R. S. (2004), *Gersgorin and his Circles*, Springer, NY.
von Neumann, J. (1929), "Zur algebra der funktionaloperatoren und theorie der normalen operatoren," *Math. Ann.*, vol. 102, pp. 370–427.
von Neumann, J. (1937), "Some matrix-inequalities and metrization of matrix-space"' *Tomsk Univ. Rev.*, vol. 1, pp. 286–300. Reprinted in Collected Works, Pergamon Press, 1962, vol. IV, pp. 205–219.
von Neumann, J. (1932), *The Mathematical Foundations of Quantum Mechanics,* translation published in 1996 of original text, Princeton University Press, NJ.

Weyl, H. (1909), "Über beschrankte quadratiche formen, deren differenz vollsteig ist.," *Rend. Circ. Mat. Palermo*, vol. 27, pp. 373–392.

Weyl, H. (1912), "Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen," *Math. Ann.*, vol. 71, pp. 441–479.

Wilkinson, J. H. (1965), *The Algebraic Eigenvalue Problem*, Oxford University Press.

Wong, Y. K. (1935), "An application of orthogonalization process to the theory of least squares," *Ann. Math. Statistics*, vol. 6, pp. 53–75.

Woodbury, M. (1950), *Inverting Modified Matrices*, Mem. Rep. 42, Statistical Research Group, Princeton University, NJ.

Zehfuss, G. (1858), "Ueber eine gewisse Determinante," *Zeitschrift für Mathematik und Physik*, vol. 3, pp. 298–301.

Zhang, F. (2005), *The Schur Complement and Its Applications*, Springer, NY.