Appendix A

Linear Algebra and Matrix Analysis Tools

A.1 INTRODUCTION

In this appendix, we provide a review of the linear algebra terms and matrix properties used in the text. For the sake of brevity, we do not present proofs for all results stated in this appendix, nor do we discuss related results not needed in the chapters. For most of the results included, however, we do provide proofs and motivation. The reader interested in finding out more about the topic of this appendix can consult the books [STEWART 1973; HORN AND JOHNSON 1985; STRANG 1988; HORN AND JOHNSON 1989; GOLUB AND VAN LOAN 1989], to which we also refer for the proofs omitted here.

A.2 RANGE SPACE, NULL SPACE, AND MATRIX RANK

Let A be an $m \times n$ matrix whose elements are complex valued in general, $A \in \mathbf{C}^{m\times n}$, and let $(\cdot)^T$ and $(\cdot)^*$ denote the *transpose* and the *conjugate transpose* operator, respectively.

Definition D1: The ***range space*** of A, also called the ***column space***, is the subspace spanned by (all linear combinations of) the columns of A:

$$\mathcal{R}(A) = \{\alpha \in \mathbf{C}^{m\times 1} | \alpha = A\beta \quad \text{for} \quad \beta \in \mathbf{C}^{n\times 1}\} \tag{A.2.1}$$

The range space of A^T is usually called the ***row space*** of A, for obvious reasons.

Definition D2: The ***null space*** of A, also called the ***kernel***, is the following subspace:

$$\mathcal{N}(A) = \{\beta \in \mathbf{C}^{n\times 1} | A\beta = 0\} \tag{A.2.2}$$

The previous definitions are all that we need to introduce the matrix rank and its basic properties. We return to the range and null subspaces in Section A.4, where we discuss the singular-value decomposition. In particular, we derive some convenient bases and useful projectors associated with the previous matrix subspaces.

Definition D3: The following are equivalent definitions of the ***rank*** of A, denoted by

$$r \triangleq \text{rank}(A)$$

(i) r is equal to the maximum number of linearly independent columns of A. The latter number is by definition the dimension of the $\mathcal{R}(A)$; hence

$$r = \dim \mathcal{R}(A) \tag{A.2.3}$$

(ii) r is equal to the maximum number of linearly independent rows of A,

$$r = \dim \mathcal{R}(A^T) = \dim \mathcal{R}(A^*) \tag{A.2.4}$$

(iii) r is the dimension of the nonzero determinant of maximum size that can be built from the elements of A.

The equivalence between the preceding Definitions (i) and (ii) is an important and pleasing result (without which one should have had to consider the row rank and column rank of a matrix separately!).

Definition D4: A is said to be

- ***Rank deficient*** whenever $r < \min(m, n)$.
- ***Full column rank*** if $r = n \leq m$.
- ***Full row rank*** if $r = m \leq n$.
- ***Nonsingular*** whenever $r = m = n$.

Result R1: Premultiplication or postmultiplication of A by a nonsingular matrix does not change the rank of A.

Proof: This fact directly follows from the definition of rank(A), because the aforementioned multiplications do not change the number of linearly independent columns (or rows) of A. ■

Result R2: Let $A \in \mathbf{C}^{m\times n}$ and $B \in \mathbf{C}^{n\times p}$ be two conformable matrices of rank r_A and r_B, respectively. Then

$$\text{rank}(AB) \leq \min(r_A,\ r_B) \tag{A.2.5}$$

Proof: We can prove the previous assertion by using the definition of the rank once again. Indeed, premultiplication of B by A cannot increase the number of linearly independent columns of B, hence $\text{rank}(AB) \leq r_B$. Similarly, postmultiplication of A by B cannot increase the number of linearly independent columns of A^T, which means that $\text{rank}(AB) \leq r_A$. ■

Result R3: Let $A \in \mathbf{C}^{m\times m}$ be given by

$$A = \sum_{k=1}^{N} x_k \, y_k^*$$

where $x_k, \; y_k \in \mathbf{C}^{m\times 1}$. Then,

$$\text{rank}(A) \leq \min(m, N)$$

Proof: A can be rewritten as

$$A = [x_1 \ldots x_N] \begin{bmatrix} y_1^* \\ \vdots \\ y_N^* \end{bmatrix}$$

so the result follows from R2. ■

Result R4: Let $A \in \mathbf{C}^{m\times n}$ with $n \leq m$, let $B \in \mathbf{C}^{n\times p}$, and let

$$\text{rank}(A) = n \tag{A.2.6}$$

Then

$$\text{rank}(AB) = \text{rank}(B) \tag{A.2.7}$$

Proof: Assumption (A.2.6) implies that A contains a nonsingular $n \times n$ submatrix, the postmultiplication of which by B gives a block of rank equal to $\text{rank}(B)$ (*cf.* R1). Hence,

$$\text{rank}(AB) \geq \text{rank}(B)$$

However, by R2, $\text{rank}(AB) \leq \text{rank}(B)$; hence, (A.2.7) follows. ■

A.3 EIGENVALUE DECOMPOSITION

Definition D5: We say that the matrix $A \in \mathbf{C}^{m\times m}$ is ***Hermitian*** if $A^* = A$. In the real-valued case, such an A is said to be ***symmetric***.

Definition D6: A matrix $U \in \mathbf{C}^{m \times m}$ is said to be ***unitary*** (***orthogonal*** if U is real valued) whenever

$$U^*U = UU^* = I$$

If $U \in \mathbf{C}^{m \times n}$, with $m > n$, is such that $U^*U = I$, then we say that U is ***semiunitary***.

Next, we present a number of definitions and results pertaining to the matrix eigenvalue decomposition (EVD), first for general matrices and then for Hermitian ones.

A.3.1 General Matrices

Definition D7: A scalar $\lambda \in \mathbf{C}$ and a (nonzero) vector $x \in \mathbf{C}^{m \times 1}$ are an ***eigenvalue*** and its associated ***eigenvector*** of a matrix $A \in \mathbf{C}^{m \times m}$ if

$$Ax = \lambda x \tag{A.3.1}$$

In particular, an eigenvalue λ is a solution of the so-called ***characteristic equation*** of A, namely,

$$|A - \lambda I| = 0 \tag{A.3.2}$$

(where $|\cdot|$ denotes determinant) and x is a vector in $\mathcal{N}(A - \lambda I)$. The pair (λ, x) is called an ***eigenpair***.

Observe that, if $\{(\lambda_i, x_i)\}_{i=1}^{p}$ are p eigenpairs of A (with $p \le m$), then we can write the defining equations $Ax_i = \lambda x_i$ $(i = 1, \ldots, p)$ in the compact form

$$AX = X\Lambda \tag{A.3.3}$$

where

$$X = [x_1 \ldots x_p]$$

and

$$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \vdots & \\ 0 & & \lambda_p \end{bmatrix}$$

Result R5: Let (λ, x) be an eigenpair of $A \in \mathbf{C}^{m \times m}$. If $B = A + \alpha I$, with $\alpha \in \mathbf{C}$, then $(\lambda + \alpha, x)$ is an eigenpair of B.

Proof: The result follows from the fact that

$$Ax = \lambda x \implies (A + \alpha I)x = (\lambda + \alpha)x. \qquad \blacksquare$$

Result R6: The matrices A and $B \triangleq Q^{-1}AQ$, where Q is any nonsingular matrix, share the same eigenvalues. (B is said to be related to A by a ***similarity transformation***.)

Proof: Indeed, the equation

$$|B - \lambda I| = |Q^{-1}(A - \lambda I)Q| = |Q^{-1}||A - \lambda I||Q| = 0$$

is equivalent to $|A - \lambda I| = 0$. ∎

In general, there is no simple relationship between the elements $\{A_{ij}\}$ of A and its eigenvalues $\{\lambda_k\}$. However, the *trace* of A, which is the sum of the diagonal elements of A, is related in a simple way to the eigenvalues, as described next.

Definition D8: The ***trace*** of a square matrix $A \in \mathbf{C}^{m\times m}$ is defined as

$$\operatorname{tr}(A) = \sum_{i=1}^{m} A_{ii} \tag{A.3.4}$$

Result R7: If $\{\lambda_i\}_{i=1}^{m}$ are the eigenvalues of $A \in \mathbf{C}^{m\times m}$, then

$$\operatorname{tr}(A) = \sum_{i=1}^{m} \lambda_i \tag{A.3.5}$$

Proof: We can write

$$|\lambda I - A| = \prod_{i=1}^{n}(\lambda - \lambda_i) \tag{A.3.6}$$

The right-hand side of (A.3.6) is a polynomial in λ whose λ^{n-1} coefficient is $\sum_{i=1}^{n} \lambda_i$. From the definition of the determinant (see, e.g., [STRANG 1988]), we find that the left-hand side of (A.3.6) is a polynomial whose λ^{n-1} coefficient is $\sum_{i=1}^{n} A_{ii} = \operatorname{tr}(A)$. This proves the result. ∎

Interestingly, although the matrix product is not commutative, the trace is invariant to commuting the factors in a matrix product, as shown next.

Result R8: Let $A \in \mathbf{C}^{m\times n}$ and $B \in \mathbf{C}^{n\times m}$. Then

$$\operatorname{tr}(AB) = \operatorname{tr}(BA) \tag{A.3.7}$$

Proof: A straightforward calculation, based on the definition of tr(·) in (A.3.4), shows that

$$\operatorname{tr}(AB) = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} B_{ji}$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{m} B_{ji} A_{ij} = \sum_{j=1}^{n} [BA]_{jj} = \operatorname{tr}(BA) \qquad \blacksquare$$

We can also prove (A.3.7) by using Result R7. Along the way, we will obtain some other useful results. First, we note the following:

Result R9: Let $A, B \in \mathbf{C}^{m \times m}$ and let $\alpha \in \mathbf{C}$. Then

$$|AB| = |A|\ |B|$$
$$|\alpha A| = \alpha^m |A|$$

Proof: The identities follow directly from the definition of the determinant; see, for example, [Strang 1988]. ■

Next we prove the following results:

Result R10: Let $A \in \mathbf{C}^{m \times n}$ and $B \in \mathbf{C}^{n \times m}$. Then

$$|I - AB| = |I - BA|. \tag{A.3.8}$$

Proof: It is straightforward to verify that

$$\begin{bmatrix} I & A \\ 0 & I \end{bmatrix} \begin{bmatrix} I & -A \\ -B & I \end{bmatrix} \begin{bmatrix} I & 0 \\ B & I \end{bmatrix} = \begin{bmatrix} I - AB & 0 \\ 0 & I \end{bmatrix} \tag{A.3.9}$$

and

$$\begin{bmatrix} I & 0 \\ B & I \end{bmatrix} \begin{bmatrix} I & -A \\ -B & I \end{bmatrix} \begin{bmatrix} I & A \\ 0 & I \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I - BA \end{bmatrix} \tag{A.3.10}$$

Because the matrices in the left-hand sides of (A.3.9) and (A.3.10) have the same determinant, equal to $\begin{vmatrix} I & -A \\ -B & I \end{vmatrix}$, it follows that the right-hand sides must also have the same determinant, which concludes the proof. ■

Result R11: Let $A \in \mathbf{C}^{m \times n}$ and $B \in \mathbf{C}^{n \times m}$. The nonzero eigenvalues of AB and of BA are identical.

Proof: Let $\lambda \neq 0$ be an eigenvalue of AB. Then,

$$0 = |AB - \lambda I| = \lambda^m |AB/\lambda - I| = \lambda^m |BA/\lambda - I| = \lambda^{m-n} |BA - \lambda I|$$

where the third equality follows from R10. Hence, λ is also an eigenvalue of BA. ■

We can now obtain R8 as a simple corollary of R11, by using the property (A.3.5) of the trace operator.

A.3.2 Hermitian Matrices

An important property of the class of Hermitian matrices, which does not necessarily hold for general matrices, is the following:

Result R12:

(i) All eigenvalues of $A = A^* \in \mathbf{C}^{m \times m}$ are ***real valued***.
(ii) The m eigenvectors of $A = A^* \in \mathbf{C}^{m \times m}$ form an ***orthonormal set***. In other words, the matrix U, whose columns are the eigenvectors of A, is ***unitary***.

It follows from (i) and (ii) and from (A.3.3) that, for a Hermitian matrix, we can write

$$AU = U\Lambda$$

where $U^*U = UU^* = I$ and the diagonal elements of Λ are real numbers. Equivalently,

$$A = U\Lambda U^* \tag{A.3.11}$$

which is the so-called eigenvalue decomposition (EVD) of $A = A^*$. The EVD of a Hermitian matrix is a special case of the singular value decomposition of a general matrix, discussed in the next section.

The following is a useful result associated with Hermitian matrices:

Result R13: Let $A = A^* \in \mathbf{C}^{m \times m}$ and let $v \in \mathbf{C}^{m \times 1}$ $(v \neq 0)$. Also, let the eigenvalues of A be arranged in a nonincreasing order:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$$

Then

$$\lambda_m \leq \frac{v^*Av}{v^*v} \leq \lambda_1 \tag{A.3.12}$$

The ratio in (A.3.12) is called the ***Rayleigh quotient***. Because this ratio is invariant under the multiplication of v by any complex number, we can rewrite (A.3.12) in the form:

$$\lambda_m \leq v^* A v \leq \lambda_1 \quad \textit{for any } v \in \mathbf{C}^{m\times 1} \textit{ with } v^* v = 1 \tag{A.3.13}$$

The equalities in (A.3.13) are evidently achieved when v is equal to the eigenvector of A associated with λ_m and λ_1, respectively.

Proof: Let the EVD of A be given by (A.3.11), and let

$$w = U^* v = \begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix}$$

We need to prove that

$$\lambda_m \leq w^* \Lambda w = \sum_{k=1}^{m} \lambda_k |w_k|^2 \leq \lambda_1$$

for any $w \in \mathbf{C}^{m\times 1}$ satisfying

$$w^* w = \sum_{k=1}^{m} |w_k|^2 = 1.$$

However, this is readily verified, as

$$\lambda_1 - \sum_{k=1}^{m} \lambda_k |w_k|^2 = \sum_{k=1}^{m} (\lambda_1 - \lambda_k)|w_k|^2 \geq 0$$

and

$$\sum_{k=1}^{m} \lambda_k |w_k|^2 - \lambda_m = \sum_{k=1}^{m} (\lambda_k - \lambda_m)|w_k|^2 \geq 0$$

and the proof is concluded. ■

The following result is an extension of R13.

Result R14: Let $V \in \mathbf{C}^{m\times n}$, with $m > n$, be a semiunitary matrix (i.e., $V^* V = I$), and let $A = A^* \in \mathbf{C}^{m\times m}$ have its eigenvalues ordered as in R13. Then

$$\sum_{k=m-n+1}^{m} \lambda_k \leq \operatorname{tr}(V^* A V) \leq \sum_{k=1}^{n} \lambda_k \tag{A.3.14}$$

where the equalities are achieved, for instance, when the columns of V are the eigenvectors of A corresponding to $(\lambda_{m-n+1}, \ldots, \lambda_m)$ and, respectively, to $(\lambda_1, \ldots, \lambda_n)$. The ratio

$$\frac{\operatorname{tr}(V^*AV)}{\operatorname{tr}(V^*V)} = \frac{\operatorname{tr}(V^*AV)}{n}$$

is sometimes called the ***extended Rayleigh quotient***.

Proof: Let

$$A = U\Lambda U^*$$

(*cf.* (A.3.11)), and let

$$S = U^*V \triangleq \begin{bmatrix} s_1^* \\ \vdots \\ s_m^* \end{bmatrix} \qquad (m \times n)$$

(hence, s_k^* is the kth row of S). By making use of the preceding notation, we can write

$$\operatorname{tr}(V^*AV) = \operatorname{tr}(V^*U\Lambda U^*V) = \operatorname{tr}(S^*\Lambda S) = \operatorname{tr}(\Lambda SS^*) = \sum_{k=1}^{m} \lambda_k c_k \tag{A.3.15}$$

where

$$c_k \triangleq s_k^* s_k, \qquad k = 1, \ldots m \tag{A.3.16}$$

Clearly,

$$c_k \geq 0, \qquad k = 1, \ldots, m \tag{A.3.17}$$

and

$$\sum_{k=1}^{m} c_k = \operatorname{tr}(SS^*) = \operatorname{tr}(S^*S) = \operatorname{tr}(V^*UU^*V) = \operatorname{tr}(V^*V) = \operatorname{tr}(I) = n \tag{A.3.18}$$

Furthermore,

$$c_k \leq 1, \qquad k = 1, \ldots, m. \tag{A.3.19}$$

To see this, let $G \in \mathbf{C}^{m \times (m-n)}$ be such that the matrix $[S \;\; G]$ is unitary; and let g_k^* denote the kth row of G. Then, by construction,

$$[s_k^* \;\; g_k^*] \begin{bmatrix} s_k \\ g_k \end{bmatrix} = c_k + g_k^* g_k = 1 \implies c_k = 1 - g_k^* g_k \leq 1$$

which is (A.3.19).

Finally, by combining (A.3.15) with (A.3.17)–(A.3.19), we can readily verify that $\mathrm{tr}(V^*AV)$ satisfies (A.3.14), where the equalities are achieved for

$$c_1 = \cdots = c_{m-n} = 0; \quad c_{m-n+1} = \cdots = c_m = 1$$

and, respectively,

$$c_1 = \cdots = c_n = 1; \quad c_{n+1} = \cdots = c_m = 0$$

These conditions on $\{c_k\}$ are satisfied if, for example, S is equal to $[0\ I]^T$ and $[I\ 0]^T$, respectively. With this observation, the proof is concluded. ■

Result R13 is clearly a special case of Result R14. The only reason for considering R13 separately is that the simpler result R13 is used more often in the text than R14.

A.4 SINGULAR VALUE DECOMPOSITION AND PROJECTION OPERATORS

For any matrix $A \in \mathbf{C}^{m\times n}$, there exist unitary matrices $U \in \mathbf{C}^{m\times m}$ and $V \in \mathbf{C}^{n\times n}$ and a diagonal matrix $\Sigma \in \mathbf{R}^{m\times n}$ with nonnegative diagonal elements, such that

$$A = U\Sigma V^* \tag{A.4.1}$$

By appropriate permutation, the diagonal elements of Σ can be arranged in a nonincreasing order:

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(m,n)}$$

The factorization (A.4.1) is called the ***singular value decomposition*** (SVD) of A, and its existence is a significant result both from a theoretical and from a practical standpoint. We reiterate that the matrices U, Σ, and V in (A.4.1) satisfy the equations

$$\begin{aligned} U^*U &= UU^* = I \quad (m\times m) \\ V^*V &= VV^* = I \quad (n\times n) \\ \Sigma_{ij} &= \begin{cases} \sigma_i \geq 0 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases} \end{aligned}$$

The following terminology is most commonly associated with the SVD:

- The ***left singular vectors*** of A are the columns of U. These singular vectors are also the eigenvectors of the matrix AA^*.
- The ***right singular vectors*** of A are the columns of V. These vectors are also the eigenvectors of the matrix A^*A.
- The ***singular values*** of A are the diagonal elements $\{\sigma_i\}$ of Σ. Note that $\{\sigma_i\}$ are the square roots of the largest $\min(m,n)$ eigenvalues of AA^* or A^*A.

- The ***singular triple*** of A is the following triple: (singular value, left singular vector, and right singular vector; σ_k, u_k, v_k), where u_k (v_k) is the kth column of $U(V)$.

If

$$\text{rank}(A) = r \leq \min(m, n)$$

then one can show that

$$\begin{cases} \sigma_k > 0, & k = 1, \ldots, r \\ \sigma_k = 0, & k = r + 1, \ldots, \min(m, n) \end{cases}$$

Hence, for a matrix of rank r, the SVD can be written as

$$A = [\, \underbrace{U_1}_{r} \quad \underbrace{U_2}_{m-r} \,] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^* \\ V_2^* \end{bmatrix} \begin{matrix} \}\, r \\ \}\, n-r \end{matrix} = U_1 \Sigma_1 V_1^* \tag{A.4.2}$$

where $\Sigma_1 \in \mathbf{R}^{r \times r}$ is nonsingular. The factorization of A in (A.4.2) has a number of important consequences.

Result R15: Consider the SVD of $A \in \mathbf{C}^{m \times n}$ in (A.4.2), where $r \leq \min(m, n)$. Then

(i) U_1 is an orthonormal basis of $\mathcal{R}(A)$;
(ii) U_2 is an orthonormal basis of $\mathcal{N}(A^*)$;
(iii) V_1 is an orthonormal basis of $\mathcal{R}(A^*)$;
(iv) V_2 is an orthonormal basis of $\mathcal{N}(A)$.

Proof: We see that (iii) and (iv) follow from the properties (i) and (ii) as applied to A^*. To prove (i) and (ii), we need to show that

$$\mathcal{R}(A) = \mathcal{R}(U_1) \tag{A.4.3}$$

and, respectively,

$$\mathcal{N}(A^*) = \mathcal{R}(U_2) \tag{A.4.4}$$

To show (A.4.3), note that

$$\begin{aligned} \alpha \in \mathcal{R}(A) &\Rightarrow \text{there exists } \beta \text{ such that } \alpha = A\beta \Rightarrow \\ &\Rightarrow \alpha = U_1(\Sigma_1 V_1^* \beta) = U_1 \gamma \Rightarrow \alpha \in \mathcal{R}(U_1) \end{aligned}$$

so $\mathcal{R}(A) \subset \mathcal{R}(U_1)$. Also,

$$\alpha \in \mathcal{R}(U_1) \Rightarrow \text{there exists } \beta \text{ such that } \alpha = U_1 \beta$$

From (A.4.2), $U_1 = AV_1\Sigma_1^{-1}$; it follows that

$$\alpha = A(V_1\Sigma_1^{-1}\beta) = A\rho \Rightarrow \alpha \in \mathcal{R}(A)$$

which shows $\mathcal{R}(U_1) \subset \mathcal{R}(A)$. Combining $\mathcal{R}(U_1) \subset \mathcal{R}(A)$ with $\mathcal{R}(A) \subset \mathcal{R}(U_1)$ gives (A.4.3). Similarly,

$$\alpha \in \mathcal{N}(A^*) \Rightarrow A^*\alpha = 0 \Rightarrow V_1\Sigma_1 U_1^*\alpha = 0 \Rightarrow \Sigma_1^{-1}V_1^*V_1\Sigma_1 U_1^*\alpha = 0 \Rightarrow U_1^*\alpha = 0$$

Now, any vector α can be written as

$$\alpha = [U_1 \ \ U_2]\begin{bmatrix} \gamma \\ \beta \end{bmatrix}$$

because $[U_1 \ \ U_2]$ is nonsingular. However, $0 = U_1^*\alpha = U_1^*U_1\gamma + U_1^*U_2\beta = \gamma$, so $\gamma = 0$, and thus $\alpha = U_2\beta$. Thus, $\mathcal{N}(A^*) \subset \mathcal{R}(U_2)$. Finally,

$$\alpha \in \mathcal{R}(U_2) \Rightarrow \text{there exists } \beta \text{ such that } \alpha = U_2\beta$$

Then

$$A^*\alpha = V_1\Sigma_1 U_1^* U_2\beta = 0 \Rightarrow \alpha \in \mathcal{N}(A^*)$$

which leads to (A.4.4). ■

This result, readily derived by using the SVD, has a number of interesting corollaries that complement the discussion on range and null subspaces in Section A.2.

Result R16: For any $A \in \mathbf{C}^{m\times n}$, the subspaces $\mathcal{R}(A)$ and $\mathcal{N}(A^*)$ are orthogonal to each other, and they together span $\mathbf{C}^m$. Consequently, we say that $\mathcal{N}(A^*)$ is the ***orthogonal complement*** of $\mathcal{R}(A)$ in $\mathbf{C}^m$, and vice versa. In particular, we have

$$\dim \mathcal{N}(A^*) = m - r \tag{A.4.5}$$

$$\dim \mathcal{N}(A) = n - r \tag{A.4.6}$$

(Recall that $\dim \mathcal{R}(A) = \dim \mathcal{R}(A^*) = r$.)

Proof: This result is a direct corollary of R15. ■

The SVD of a matrix also provides a convenient representation for the projectors onto the range and null spaces of A and A^*.

Definition D9: Let $y \in \mathbf{C}^{m\times 1}$ be an arbitrary vector. By definition, the ***orthogonal projector*** onto $\mathcal{R}(A)$ is the matrix Π, which is such that (i) $\mathcal{R}(\Pi) = \mathcal{R}(A)$ and (ii) the Euclidean distance between y and $\Pi y \in \mathcal{R}(A)$ is minimum:

$$\|y - \Pi y\|^2 = \min \quad \text{over } \mathcal{R}(A)$$

Hereafter, $\|x\| = \sqrt{x^* x}$ denotes the ***Euclidean vector norm***.

Result R17: Let $A \in \mathbf{C}^{m\times n}$. The orthogonal projector onto $\mathcal{R}(A)$ is given by

$$\Pi = U_1 U_1^* \tag{A.4.7}$$

whereas the orthogonal projector onto $\mathcal{N}(A^*)$ is

$$\Pi^{\perp} = I - U_1 U_1^* = U_2 U_2^* \tag{A.4.8}$$

Proof: Let $y \in \mathbf{C}^{m\times 1}$ be an arbitrary vector. As $\mathcal{R}(A) = \mathcal{R}(U_1)$, according to R15, we can find the vector in $\mathcal{R}(A)$ that is of minimal distance from y by solving the problem

$$\min_{\beta} \|y - U_1 \beta\|^2 \tag{A.4.9}$$

Because

$$\begin{aligned}\|y - U_1\beta\|^2 &= (\beta^* - y^* U_1)(\beta - U_1^* y) + y^*(I - U_1 U_1^*) y \\ &= \|\beta - U_1^* y\|^2 + \|U_2^* y\|^2\end{aligned}$$

it readily follows that the solution to the minimization problem (A.4.9) is given by $\beta = U_1^* y$. Hence, the vector $U_1 U_1^* y$ is the orthogonal projection of y onto $\mathcal{R}(A)$, and the minimum distance from y to $\mathcal{R}(A)$ is $\|U_2^* y\|$. This proves (A.4.7). Then (A.4.8) follows immediately from (A.4.7) and the fact that $\mathcal{N}(A^*) = \mathcal{R}(U_2)$. ■

Note, for instance, that, for the projection of y onto $\mathcal{R}(A)$, the error vector is $y - U_1 U_1^* y = U_2 U_2^* y$, which is in $\mathcal{R}(U_2)$ and therefore is orthogonal to $\mathcal{R}(A)$ by R15. For this reason, Π is given the name "orthogonal projector" in D9 and R17.

As an aside, we remark that the orthogonal projectors in (A.4.7) and (A.4.8) are *idempotent matrices*; see the next definition.

Definition D10: The matrix $A \in \mathbf{C}^{m\times m}$ is ***idempotent*** if

$$A^2 = A \tag{A.4.10}$$

Furthermore, observe, by making use of R11, that the idempotent matrix in (A.4.7), for example, has r eigenvalues equal to 1 and $(m - r)$ eigenvalues equal to 0. This is a general property of idempotent matrices: their eigenvalues are either 0 or 1.

Finally, we present a result that, even alone, would be enough to make the SVD an essential matrix-analysis tool.

Result R18: Let $A \in \mathbf{C}^{m \times n}$, with elements A_{ij}. Let the SVD of A (with the singular values arranged in a nonincreasing order) be given by

$$A = [\underbrace{U_1}_{p} \quad \underbrace{U_2}_{m-p}] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^* \\ V_2^* \end{bmatrix} \begin{matrix} \}\, p \\ \}\, n-p \end{matrix} \tag{A.4.11}$$

where $p \leq \min(m, n)$ is an integer. Let

$$\|A\|^2 = \operatorname{tr}(A^*A) = \sum_{i=1}^{m} \sum_{j=1}^{n} |A_{ij}|^2 = \sum_{k=1}^{\min(m,n)} \sigma_k^2 \tag{A.4.12}$$

denote the square of the so-called ***Frobenius norm***. Then the ***best rank-p approximant*** of A in the Frobenius-norm metric, that is, the solution to

$$\min_{B} \ \|A - B\|^2 \quad \text{subject to } \operatorname{rank}(B) = p \ , \tag{A.4.13}$$

is given by

$$B_0 = U_1 \Sigma_1 V_1^* \tag{A.4.14}$$

Furthermore, B_0 is the unique solution to the approximation problem (A.4.13) if and only if $\sigma_p > \sigma_{p+1}$.

Proof: It follows from R4 and (A.4.2) that we can parameterize B in (A.4.13) as

$$B = CD^* \tag{A.4.15}$$

where $C \in \mathbf{C}^{m \times p}$ and $D \in \mathbf{C}^{n \times p}$ are full-column-rank matrices. The previous parameterization of B is of course nonunique, but, as we will see, this fact does not introduce any problem. By making use of (A.4.15), we can rewrite the problem (A.4.13) in the following form:

$$\min_{C,D} \ \|A - CD^*\|^2 \quad \operatorname{rank}(C) = \operatorname{rank}(D) = p \tag{A.4.16}$$

The reparameterized problem is essentially constraint free. Indeed, the full-column-rank condition that must be satisfied by C and D can be easily handled.

First, we minimize (A.4.16) with respect to D, for a given C. To that end, observe that

$$\begin{aligned} \|A - CD^*\|^2 = \operatorname{tr}\{&[D - A^*C(C^*C)^{-1}](C^*C)[D^* - (C^*C)^{-1}C^*A] \\ &+ A^*[I - C(C^*C)^{-1}C^*]A\} \end{aligned} \tag{A.4.17}$$

By result (iii) in Definition D11 in the next section, the matrix $[D - A^*C(C^*C)^{-1}] \cdot (C^*C)[D^* - (C^*C)^{-1}C^*A]$ is positive semidefinite for any D. This observation implies that (A.4.17) is minimized with respect to D for

$$D_0 = A^*C(C^*C)^{-1} \tag{A.4.18}$$

and that the corresponding minimum value of (A.4.17) is given by

$$\operatorname{tr}\{A^*[I - C(C^*C)^{-1}C^*]A\} \tag{A.4.19}$$

Next, we minimize (A.4.19) with respect to C. Let $S \in \mathbf{C}^{m \times p}$ denote an orthogonal basis of $\mathcal{R}(C)$—that is, $S^*S = I$ and

$$S = C\Gamma$$

for some nonsingular $p \times p$ matrix Γ. It is then straightforward to verify that

$$I - C(C^*C)^{-1}C^* = I - SS^* \tag{A.4.20}$$

By combining (A.4.19) and (A.4.20), we can restate the problem of minimizing (A.4.19) with respect to C as

$$\max_{S;\, S^*S=I} \operatorname{tr}[S^*(AA^*)S] \tag{A.4.21}$$

The solution to (A.4.21) follows from R14; the maximizing S is given by

$$S_0 = U_1$$

which yields

$$C_0 = U_1\Gamma^{-1} \tag{A.4.22}$$

It follows that

$$\begin{aligned} B_0 &= C_0D_0^* = C_0(C_0^*C_0)^{-1}C_0^*A = S_0S_0^*A \\ &= U_1U_1^*(U_1\Sigma_1V_1^* + U_2\Sigma_2V_2^*) \\ &= U_1\Sigma_1V_1^*. \end{aligned}$$

Furthermore, we observe that the minimum value of the Frobenius distance in (A.4.13) is given by

$$\|A - B_0\|^2 = \|U_2\Sigma_2V_2^*\|^2 = \sum_{k=p+1}^{\min(m,n)} \sigma_k^2$$

If $\sigma_p > \sigma_{p+1}$, then the best rank-p approximant B_0 is unique; otherwise, it is not unique. Indeed, whenever $\sigma_p = \sigma_{p+1}$, we can obtain B_0 by using either the singular vectors associated with σ_p or those corresponding to σ_{p+1}; each alternative choice generally leads to a different solution.

■

A.5 POSITIVE (SEMI)DEFINITE MATRICES

Let $A = A^* \in \mathbf{C}^{m\times m}$ be a Hermitian matrix, and let $\{\lambda_k\}_{k=1}^m$ denote its eigenvalues.

Definition D11: We say that A is ***positive semidefinite*** (psd) or ***positive definite*** (pd) if any of the following equivalent conditions holds true:

(i) $\lambda_k \geq 0$ ($\lambda_k > 0$ for pd) for $k = 1, \ldots, m$.
(ii) $\alpha^* A\alpha \geq 0$ ($\alpha^* A\alpha > 0$ for pd) for any nonzero vector $\alpha \in \mathbf{C}^{m\times 1}$
(iii) There exists a matrix C such that

$$A = CC^* \tag{A.5.1}$$

(with $\text{rank}(C) = m$ for pd)
(iv) $|A(i_1, \ldots, i_k)| \geq 0$ (> 0 for pd) for all $k = 1, \ldots, m$ and all indices $i_1, \ldots, i_k \in [1, m]$, where $A(i_1, \ldots, i_k)$ is the submatrix formed from A by eliminating the $i_1, \ldots, i_k$ rows and columns of A. ($A(i_1, \ldots, i_k)$ is called a ***principal submatrix*** of A). The condition for A to be positive definite can be simplified to requiring that $|A(k+1, \ldots, m)| > 0$ (for $k = 1, \ldots, m-1$) and $|A| > 0$. ($A(k+1, \ldots, m)$ is called a ***leading submatrix*** of A).

The notation $A > 0$ ($A \geq 0$) is commonly used to denote that A is pd (psd).

Of the previous defining conditions, (iv) is apparently the most involved. The necessity of (iv) can be proven as follows: Let α be a vector in $\mathbf{C}^m$ with zeroes at the positions $\{i_1, \ldots, i_k\}$ and arbitrary elements elsewhere. Then, by using (ii), we readily see that $A \geq 0$ (> 0) implies $A(i_1, \ldots, i_k) \geq 0$ (> 0), which, in turn, implies (iv) by making use of (i) and the fact that the determinant of a matrix equals the product of its eigenvalues. The sufficiency of (iv) is shown in [Strang 1988].

The equivalence of the remaining conditions, (i), (ii), and (iii), is easily proven by making use of the EVD of A: $A = U\Lambda U^*$. To show that (i) $\Leftrightarrow$ (ii), assume first that (i) holds and let $\beta = U^*\alpha$. Then

$$\alpha^* A\alpha = \beta^*\Lambda\beta = \sum_{k=1}^{m} \lambda_k |\beta_k|^2 \geq 0 \tag{A.5.2}$$

and hence, (ii) holds as well. Conversely, because U is invertible, it follows from (A.5.2) that (ii) can hold, only if (i) holds; indeed, if (i) does not hold, one can choose β to make (A.5.2) negative; thus there exists an $\alpha = U\beta$ such that $\alpha^* A\alpha < 0$, which contradicts the assumption that (ii) holds. Consequently, (i) and (ii) are equivalent. To show that (iii) $\Rightarrow$ (ii), note that

$$\alpha^* A\alpha = \alpha^* CC^*\alpha = \|C^*\alpha\|^2 \geq 0$$

and thus (ii) holds as well. Because (iii) $\Rightarrow$ (ii) and (ii) $\Rightarrow$ (i), we have (iii) $\Rightarrow$ (i). To show that (i) $\Rightarrow$ (iii), we assume (i) and write

$$A = U\Lambda U^* = (U\Lambda^{1/2}\Lambda^{1/2}U^*) = (U\Lambda^{1/2}U^*)(U\Lambda^{1/2}U^*) \triangleq CC^* \tag{A.5.3}$$

and thus (iii) is also satisfied. In (A.5.3), $\Lambda^{1/2}$ is a diagonal matrix whose diagonal elements are equal to $\{\lambda_k^{1/2}\}$. In other words, $\Lambda^{1/2}$ is the "square root" of Λ.

In a general context, the square root of a positive semidefinite matrix is defined as follows:

Definition D12: Let $A = A^*$ be a positive semidefinite matrix. Then any matrix C that satisfies

$$A = CC^* \tag{A.5.4}$$

is called a ***square root*** of A. Sometimes such a C is denoted by $A^{1/2}$.

If C is a square root of A, then so is CB for any unitary matrix B; hence, a given positive semidefinite matrix has an infinite number of square roots. Two often-used particular choices for square roots are

(i) *Hermitian square root*: $C = C^*$. In this case, we can write (A.5.4) as $A = C^2$. Note that we have already obtained such a square root of A in (A.5.3):

$$C = U\Lambda^{1/2}U^* \tag{A.5.5}$$

If C is also constrained to be positive semidefinite ($C \geq 0$) then the Hermitian square root is unique.

(ii) *Cholesky factor*. If C is lower triangular with nonnegative diagonal elements, then C is called the *Cholesky factor* of A. In computational exercises, the triangular form of the square-root matrix is often preferred to other forms. If A is positive definite, the Cholesky factor is unique.

We also note that equation (A.5.4) implies that A and C *have the same rank and the same range space*. This follows easily, for example, from inserting the SVD of C into (A.5.4).

Next, we prove three specialized results on positive semidefinite matrices required in Section 2.5 and in Appendix B.

Result R19: Let $A \in \mathbf{C}^{m\times m}$ and $B \in \mathbf{C}^{m\times m}$ be positive semidefinite matrices. Then the matrix $A \odot B$ is also positive semidefinite, where $\odot$ denotes the ***Hadamard matrix product*** (also called ***elementwise multiplication***: $[A \odot B]_{ij} = A_{ij}B_{ij}$).

Proof: Because B is positive semidefinite, it can be written as $B = CC^*$ for some matrix $C \in \mathbf{C}^{m\times m}$. Let c_k^* denote the kth row of C. Then,

$$[A \odot B]_{ij} = A_{ij}B_{ij} = A_{ij}\; c_i^* c_j$$

and, hence, for any $\alpha \in \mathbf{C}^{m\times 1}$,

$$\alpha^*(A \odot B)\alpha = \sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i^* A_{ij} c_i^* c_j \alpha_j \tag{A.5.6}$$

By letting $\{c_{jk}\}_{k=1}^{m}$ denote the elements of the vector c_j, we can rewrite (A.5.6) as

$$\alpha^*(A \odot B)\alpha = \sum_{k=1}^{m} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i^* c_{ik}^* A_{ij} \alpha_j c_{jk} = \sum_{k=1}^{m} \beta_k^* A \beta_k \tag{A.5.7}$$

where

$$\beta_k \triangleq [\alpha_1 c_{1k} \cdots \alpha_m c_{mk}]^T$$

A is positive semidefinite by assumption, so $\beta_k^* A \beta_k \geq 0$ for each k, and it follows from (A.5.7) that $A \odot B$ must be positive semidefinite as well. ■

Result R20: Let $A \in \mathbf{C}^{m \times m}$ and $B \in \mathbf{C}^{m \times m}$ be Hermitian matrices. Assume that B is nonsingular and that the partitioned matrix

$$\begin{bmatrix} A & I \\ I & B \end{bmatrix}$$

is positive semidefinite. Then the matrix $(A - B^{-1})$ is also positive semidefinite:

$$A \geq B^{-1}$$

Proof: By Definition D11, part (ii),

$$\begin{bmatrix} \alpha_1^* & \alpha_2^* \end{bmatrix} \begin{bmatrix} A & I \\ I & B \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \geq 0 \tag{A.5.8}$$

for any vectors $\alpha_1, \alpha_2 \in \mathbf{C}^{m \times 1}$. Let

$$\alpha_2 = -B^{-1}\alpha_1$$

Then (A.5.8) becomes

$$\alpha_1^*(A - B^{-1})\alpha_1 \geq 0$$

This inequality must hold for any $\alpha_1 \in \mathbf{C}^{m \times 1}$, and so the proof is concluded. ■

Result R21: Let $C \in \mathbf{C}^{m \times m}$ be a (Hermitian) positive definite matrix depending on a real-valued parameter α. Assume that C is a differentiable function of α. Then

$$\frac{\partial}{\partial \alpha}[\ln |C|] = \operatorname{tr}\left[C^{-1} \frac{\partial C}{\partial \alpha}\right]$$

Proof: Let $\{\lambda_i\} \in \mathbf{R}$ ($i = 1, \ldots, m$) denote the eigenvalues of C. Then

$$\begin{aligned}
\frac{\partial}{\partial \alpha}[\ln |C|] = \frac{\partial}{\partial \alpha}\left[\ln \prod_{k=1}^{m} \lambda_k\right] &= \sum_{k=1}^{m} \frac{\partial}{\partial \alpha}(\ln \lambda_k) \\
&= \sum_{k=1}^{m} \frac{1}{\lambda_k}\frac{\partial \lambda_k}{\partial \alpha} \\
&= \operatorname{tr}\left[\Lambda^{-1}\frac{\partial \Lambda}{\partial \alpha}\right]
\end{aligned}$$

where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_m)$. Let Q be a unitary matrix such that $Q^*\Lambda Q = C$ (which is the EVD of C). Since Q is unitary, $Q^*Q = I$, we obtain

$$\frac{\partial Q^*}{\partial \alpha}Q + Q^*\frac{\partial Q}{\partial \alpha} = 0$$

Thus, we get

$$\begin{aligned}
\operatorname{tr}\left[\Lambda^{-1}\frac{\partial \Lambda}{\partial \alpha}\right] &= \operatorname{tr}\left[\left(Q^*\Lambda^{-1}Q\right)\left(Q^*\frac{\partial \Lambda}{\partial \alpha}Q\right)\right] \\
&= \operatorname{tr}\left[C^{-1}\left(\frac{\partial}{\partial \alpha}\left(Q^*\Lambda Q\right) - \frac{\partial Q^*}{\partial \alpha}\Lambda Q - Q^*\Lambda\frac{\partial Q}{\partial \alpha}\right)\right] \\
&= \operatorname{tr}\left[C^{-1}\frac{\partial C}{\partial \alpha}\right] - \operatorname{tr}\left[Q^*\Lambda^{-1}Q\left(\frac{\partial Q^*}{\partial \alpha}\Lambda Q + Q^*\Lambda\frac{\partial Q}{\partial \alpha}\right)\right] \\
&= \operatorname{tr}\left[C^{-1}\frac{\partial C}{\partial \alpha}\right] - \operatorname{tr}\left[\frac{\partial Q^*}{\partial \alpha}Q + Q^*\frac{\partial Q}{\partial \alpha}\right] \\
&= \operatorname{tr}\left[C^{-1}\frac{\partial C}{\partial \alpha}\right]
\end{aligned}$$

which is the result stated. ■

Finally, we make use of a simple property of positive semidefinite matrices to prove the ***Cauchy–Schwartz inequality*** for vectors and for functions.

Result R22 (Cauchy–Schwartz inequality for vectors): Let $x,\ y \in \mathbf{C}^{m\times 1}$. Then

$$|x^*y|^2 \le \|x\|^2\ \|y\|^2 \tag{A.5.9}$$

where $|\cdot|$ denotes the modulus of a possibly complex-valued number, and $\|\cdot\|$ denotes the Euclidean vector norm ($\|x\|^2 = x^*x$). Equality in (A.5.9) is achieved if and only if x is proportional to y.

Proof: The (2×2) matrix

$$\begin{bmatrix} \|x\|^2 & x^*y \\ y^*x & \|y\|^2 \end{bmatrix} = \begin{bmatrix} x^* \\ y^* \end{bmatrix} \begin{bmatrix} x & y \end{bmatrix} \tag{A.5.10}$$

is positive semidefinite because condition (iii) in D11 is satisfied. It follows from condition (iv) in D11 that the determinant of the preceding matrix must be nonnegative; in other words,

$$\|x\|^2 \, \|y\|^2 - |x^*y|^2 \geq 0$$

which gives (A.5.9). Equality in (A.5.9) holds if and only if the determinant of (A.5.10) is equal to zero. The latter condition is equivalent to requiring that x be proportional to y. (*Cf.* D3: The columns of the matrix $[x \ y]$ will then be linearly dependent.) ■

Result R23 (Cauchy–Schwartz inequality for functions): Let $f(x)$ and $g(x)$ be two complex-valued functions defined for a real-valued argument x. Then, assuming that the needed integrals exist,

$$\left| \int_I f(x) g^*(x) dx \right|^2 \leq \left[\int_I |f(x)|^2 dx \right] \left[\int_I |g(x)|^2 dx \right]$$

where $I \subset \mathbf{R}$ is an integration interval. The inequality above becomes an equality if and only if $f(x)$ is proportional to $g(x)$ on I.

Proof: The matrix

$$\int_I \begin{bmatrix} f(x) \\ g(x) \end{bmatrix} \begin{bmatrix} f^*(x) & g^*(x) \end{bmatrix} dx$$

is seen to be positive semidefinite (because the integrand is a positive semidefinite matrix for every $x \in I$). Hence, the stated result follows from the type of argument used in the proof of Result R22. ■

A.6 MATRICES WITH SPECIAL STRUCTURE

In this section, we consider several types of matrices with a special structure, for which we prove some basic properties used in the text.

Definition D13: A matrix $A \in \mathbf{C}^{m \times n}$ is called ***Vandermonde*** if it has the structure

$$A = \begin{bmatrix} 1 & \cdots & 1 \\ z_1 & & z_n \\ \vdots & & \vdots \\ z_1^{m-1} & \cdots & z_n^{m-1} \end{bmatrix} \tag{A.6.1}$$

where $z_k \in \mathbf{C}$ are usually assumed to be distinct.

Result R24: Consider the matrix A in (A.6.1) with $z_k \neq z_p$ for $k, p = 1, \dots, n$ and $k \neq p$. Also let $m \geq n$ and assume that $z_k \neq 0$ for all k. Then any n consecutive rows of A are linearly independent.

Proof: To prove the assertion, it is sufficient to show that the following $n \times n$ Vandermonde matrix is nonsingular:

$$\bar{A} = \begin{bmatrix} 1 & \cdots & 1 \\ z_1 & & z_n \\ \vdots & & \vdots \\ z_1^{n-1} & \cdots & z_n^{n-1} \end{bmatrix}$$

Let $\beta = [\beta_0 \cdots \beta_{n-1}]^* \neq 0$. The equation $\beta^* \bar{A} = 0$ is equivalent to

$$\beta_0 + \beta_1 z + \cdots + \beta_{n-1} z^{n-1} = 0 \quad \text{at} \quad z = z_k \quad (k = 1, \dots, n) \tag{A.6.2}$$

However, (A.6.2) is impossible; an $(n-1)$-degree polynomial cannot have n zeroes. Hence, $\bar{A}$ has full rank. ■

Definition D14: A matrix $A \in \mathbf{C}^{m \times n}$ is called

- ***Toeplitz*** when A_{ij} is a function of $i - j$ only.
- ***Hankel*** when A_{ij} is a function of $i + j$ only.

Observe that a Toeplitz matrix has the same element along each diagonal, whereas a Hankel matrix has identical elements on each of the antidiagonals.

Result R25: The eigenvectors of a symmetric Toeplitz matrix $A \in \mathbf{R}^{m \times m}$ are either symmetric or skew symmetric. More precisely, if J denotes the exchange (or reversal) matrix

$$J = \begin{bmatrix} 0 & & 1 \\ & \iddots & \\ 1 & & 0 \end{bmatrix}$$

and if x is an eigenvector of A, then either $x = Jx$ or $x = -Jx$.

Proof: By the property (3.5.3) proven in Section 3.5, A satisfies

$$AJx = JAx$$

or, equivalently,

$$(JAJ)x = Ax$$

for any $x \in \mathbf{C}^{m \times 1}$. Hence, we must have

$$JAJ = A \tag{A.6.3}$$

Let (λ, x) denote an eigenpair of A:

$$Ax = \lambda x \tag{A.6.4}$$

Combining (A.6.3) and (A.6.4) yields

$$\lambda Jx = JAx = J(JAJ)x = A(Jx) \tag{A.6.5}$$

Because the eigenvectors of a symmetric matrix are unique modulo multiplication by a scalar, it follows from (A.6.5) that

$$x = \alpha Jx \quad \text{for some } \alpha \in \mathbf{R}$$

As x (and, hence, Jx) must have unit norm, α must satisfy $\alpha^2 = 1 \Rightarrow \alpha = \pm 1$; thus, either $x = Jx$ (x is symmetric) or $x = -Jx$ (x is skew symmetric). ■

One can show that, for m even, the number of symmetric eigenvectors is $m/2$, as is the number of skew-symmetric eigenvectors; for m odd, the number of symmetric eigenvectors is $(m+1)/2$ and the number of skew-symmetric eigenvectors is $(m-1)/2$. (See [Cantoni and Butler 1976].)

For many additional results on Toeplitz matrices, the reader can consult [Iohvidov 1982; Böttcher and Silbermann 1983].

A.7 MATRIX INVERSION LEMMAS

The following formulas for *the inverse of a partitioned matrix* are used in the text:

Result R26: Let $A \in \mathbf{C}^{m\times m}$, $B \in \mathbf{C}^{n\times n}$, $C \in \mathbf{C}^{m\times n}$ and $D \in \mathbf{C}^{n\times m}$. Then, provided that the appropriate matrix inverses exist,

$$\begin{aligned}
\begin{bmatrix} A & C \\ D & B \end{bmatrix}^{-1} &= \begin{bmatrix} I \\ 0 \end{bmatrix} A^{-1} \begin{bmatrix} I & 0 \end{bmatrix} + \begin{bmatrix} -A^{-1}C \\ I \end{bmatrix} (B - DA^{-1}C)^{-1} [-DA^{-1} \;\; I] \\
&= \begin{bmatrix} 0 \\ I \end{bmatrix} B^{-1} \begin{bmatrix} 0 & I \end{bmatrix} + \begin{bmatrix} I \\ -B^{-1}D \end{bmatrix} (A - CB^{-1}D)^{-1} [I \;\; -CB^{-1}]
\end{aligned}$$

Proof: By direct verification. ■

By equating the top-left blocks in these two equations, we obtain the so-called *Matrix Inversion Lemma*:

Result R27 ***Matrix Inversion Lemma*****:** Let A, B, C, and D be as in R26. Then, assuming that the matrix inverses exist,

$$(A - CB^{-1}D)^{-1} = A^{-1} + A^{-1}C(B - DA^{-1}C)^{-1}DA^{-1}$$

A.8 SYSTEMS OF LINEAR EQUATIONS

Let $A \in \mathbf{C}^{m \times n}$, $B \in \mathbf{C}^{m \times p}$, and $X \in \mathbf{C}^{n \times p}$. A general *system of linear equations* in X can be written as

$$AX = B \tag{A.8.1}$$

where A and B are given and X is the unknown matrix. The special case of (A.8.1) corresponding to $p = 1$ (for which X and B are vectors) is perhaps the most common one in applications. For the sake of generality, we consider the system (A.8.1) with $p \geq 1$. (The ESPRIT system of equations encountered in Section 4.7 is of the form of (A.8.1) with $p > 1$.) We say that (A.8.1) is ***exactly determined*** whenever $m = n$, ***overdetermined*** if $m > n$ and ***underdetermined*** if $m < n$. In the following discussion, we first address the case where (A.8.1) has an exact solution and then the cases where (A.8.1) cannot be exactly satisfied.

A.8.1 Consistent Systems

Result R28: The linear system (A.8.1) is ***consistent***, that is it admits an exact solution X, if and only if $\mathcal{R}(B) \subset \mathcal{R}(A)$ or equivalently

$$\operatorname{rank}([A \; B]) = \operatorname{rank}(A) \tag{A.8.2}$$

Proof: The result is readily shown by the use of rank and range properties. ■

Result R29: Let X_0 be a particular solution to (A.8.1). Then *the set of all solutions* to (A.8.1) is given by

$$X = X_0 + \Delta \tag{A.8.3}$$

where $\Delta \in \mathbf{C}^{n \times p}$ is any matrix whose columns are in $\mathcal{N}(A)$.

Proof: Obviously, (A.8.3) satisfies (A.8.1). To show that no solution outside the set (A.8.3) exists, let $\Omega \in \mathbf{C}^{n \times p}$ be a matrix whose columns do not all belong to $\mathcal{N}(A)$. Then $A\Omega \neq 0$ and

$$A(X_0 + \Delta + \Omega) = A\Omega + B \neq B$$

and hence, $X_0 + \Delta + \Omega$ is not a solution to $AX = B$. ■

Result R30: The system of linear equations (A.8.1) has a *unique solution* if and only if (A.8.2) holds and A has full column rank:

$$\operatorname{rank}(A) = n \leq m \tag{A.8.4}$$

Proof: The assertion follows from R28 and R29. ■

Next, let us assume that (A.8.1) is consistent but A does *not* satisfy (A.8.4) (hence, $\dim \mathcal{N}(A) \geq 1$). Then, according to R29, there are infinitely many solutions. In what follows, we obtain that (unique) solution X_0 that has *minimum norm.*

Result R31: Consider a linear system that satisfies the consistency condition in (A.8.2). Let A have rank $r \leq \min(m, n)$, and let

$$A = [\,\underbrace{U_1}_{r} \quad \underbrace{U_2}_{m-r}\,] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^* \\ V_2^* \end{bmatrix} \begin{matrix} \}\, r \\ \}\, n-r \end{matrix} = U_1 \Sigma_1 V_1^*$$

denote the SVD of A. (Here Σ_1 is nonsingular, *cf.* the discussion in Section A.4). Then

$$X_0 = V_1 \Sigma_1^{-1} U_1^* B \tag{A.8.5}$$

is the ***minimum-Frobenius-norm solution*** of (A.8.1), in the sense that

$$\|X_0\|^2 < \|X\|^2 \tag{A.8.6}$$

for any other solution $X \neq X_0$.

Proof: First, we verify that X_0 satisfies (A.8.1). We have

$$AX_0 = U_1 U_1^* B \tag{A.8.7}$$

In (A.8.7), $U_1 U_1^*$ is the orthogonal projector onto $\mathcal{R}(A)$ (*cf.* R17). Because B must belong to $\mathcal{R}(A)$ (see R28), we conclude that $U_1 U_1^* B = B$ and, hence, that X_0 is indeed a solution.

Next, we note that, according to R15,

$$\mathcal{N}(A) = \mathcal{R}(V_2)$$

Consequently, the general solution (A.8.3) can be written (*cf.* R29) as

$$X = X_0 + V_2 Q\ ; \qquad Q \in \mathbf{C}^{(n-r)\times p}$$

from which we obtain

$$\begin{aligned} \|X\|^2 &= \operatorname{tr}[(X_0^* + Q^* V_2^*)(X_0 + V_2 Q)] \\ &= \|X_0\|^2 + \|V_2 Q\|^2 > \|X_0\|^2 \quad \text{for } X \neq X_0 \end{aligned}$$

■

Definition D15: The matrix

$$A^\dagger \triangleq V_1 \Sigma_1^{-1} U_1^* \tag{A.8.8}$$

in (A.8.5) is the so-called ***Moore–Penrose pseudoinverse*** (or ***generalized inverse***) of A.

It can be shown that $A^\dagger$ is the unique solution to the following set of equations:

$$\begin{cases} AA^\dagger A = A \\ A^\dagger A A^\dagger = A^\dagger \\ A^\dagger A \text{ and } AA^\dagger \text{ are Hermitian} \end{cases}$$

Evidently, whenever A is square and nonsingular, we have $A^\dagger = A^{-1}$; this observation partly motivates the name "generalized inverse" (or "pseudoinverse") given to $A^\dagger$ in the general case.

The computation of a solution to (A.8.1), whenever one exists, is an important issue, which we address briefly in what follows. We begin by noting that, in the general case, there is no computer algorithm that can compute a solution to (A.8.1) *exactly* (i.e., without any numerical errors). In effect, the best we can hope for is to compute the exact solution to a slightly perturbed (fictitious) system of linear equations, given by

$$(A + \Delta_A)(X + \Delta_X) = B + \Delta_B \tag{A.8.9}$$

where Δ_A and Δ_B are small perturbation terms, the magnitude of which depends on the algorithm and the length of the computer word, and where Δ_X is the solution perturbation induced. An algorithm which, when applied to (A.8.1), provides a solution to (A.8.9) corresponding to perturbation terms (Δ_A, Δ_B) whose magnitude is of the order afforded by the "machine epsilon" is said to be *numerically stable*. Now, assuming that (A.8.1) has a unique solution (and, hence, that A satisfies (A.8.4)), one can show that the perturbations in A and B in (A.8.9) are retrieved in Δ_X multiplied by a proportionality factor given by

$$\text{cond}(A) = \sigma_1/\sigma_n \tag{A.8.10}$$

where σ_1 and σ_n are the largest and smallest singular values of A, respectively, and where "cond" is short for "condition." The system (A.8.1) is said to be *well conditioned* if the corresponding ratio (A.8.10) is "small" (that is, not much larger than 1). The ratio in (A.8.10) is called the *condition number* of the matrix A and is an important parameter of a given system of linear equations. Note, from the previous discussion, that even a numerically stable algorithm (i.e., one that induces quite small Δ_A and Δ_B) could yield an inaccurate solution X when applied to an ill-conditioned system of linear equations (i.e., a system with a very large cond(A)). For more details on the topic of this paragraph, including specific algorithms for solving linear systems, we refer the reader to [Stewart 1973; Golub and Van Loan 1989].

A.8.2 Inconsistent Systems

The systems of linear equations that appear in applications (such as those in this book) are quite often perturbed versions of a "nominal system," and usually they do *not* admit any exact solution. Such systems are said to be *inconsistent*, and frequently they are overdetermined and have a matrix A that has full column rank:

$$\text{rank}(A) = n \le m \tag{A.8.11}$$

In what follows, we present two approaches to obtaining an approximate solution to an inconsistent system of linear equations

$$AX \simeq B \tag{A.8.12}$$

under the condition (A.8.11).

Definition D16: The ***least squares*** (LS) approximate solution to (A.8.12) is given by the minimizer X_{LS} of the following criterion:

$$\|AX - B\|^2$$

Equivalently, X_{LS} can be defined as follows: Obtain the minimal perturbation Δ_B that makes the system (A.8.12) consistent—that is,

$$\min \ \|\Delta_B\|^2 \quad \text{subject to} \quad AX = B + \Delta_B \tag{A.8.13}$$

Then derive X_{LS} by solving the system in (A.8.13) corresponding to the optimal perturbation Δ_B.

The LS solution introduced above can be obtained in several ways. A simple way is as follows:

Result R32: The LS solution to (A.8.12) is given by

$$X_{LS} = (A^*A)^{-1}A^*B \tag{A.8.14}$$

The inverse matrix in this equation exists, in view of (A.8.11).

Proof: The matrix B_0 that makes the system consistent and is of minimal distance (in the Frobenius-norm metric) from B is given by the orthogonal projection of (the columns of) B onto $\mathcal{R}(A)$:

$$B_0 = A(A^*A)^{-1}A^*B \tag{A.8.15}$$

To motivate (A.8.15) by using only the results proven so far in this appendix, we digress from the main proof and let U_1 denote an orthogonal basis of $\mathcal{R}(A)$. Then R17 implies that $B_0 = U_1U_1^*B$. However, U_1 and A span the same subspace; hence, they must be related to one another by a nonsingular linear transformation: $U_1 = AQ$ ($|Q| \neq 0$). It follows from this observation that $U_1U_1^* = AQQ^*A^*$ and also that $Q^*A^*AQ = I$, which lead to the following projector formula: $U_1U_1^* = A(A^*A)^{-1}A^*$ (as used in (A.8.15)).

Next, we return to the proof of (A.8.14). The unique solution to

$$AX - B_0 = A[X - (A^*A)^{-1}A^*B]$$

is obviously (A.8.14), because $\dim \mathcal{N}(A) = 0$ by assumption. ■

The *LS* solution X_{LS} can be computed by means of the SVD of the $m \times n$ matrix A. The X_{LS} can, however, be obtained in a computationally more efficient way, as is briefly described below. Note that X_{LS} should *not* be computed by directly evaluating the formula in (A.8.14) as it stands. Briefly stated, the reason is as follows: Recall, from (A.8.10), that the condition number of A is given by

$$\text{cond}(A) = \sigma_1/\sigma_n \tag{A.8.16}$$

(Note that $\sigma_n \neq 0$ under (A.8.11).) When working directly on A, we find that the numerical errors made in the computation of X_{LS} can be shown to be proportional to (A.8.16). However, in (A.8.14), one would need to invert the matrix A^*A, whose condition number is

$$\text{cond}(A^*A) = \sigma_1^2/\sigma_n^2 = [\text{cond}(A)]^2 \tag{A.8.17}$$

Working with (A^*A) would therefore lead to much larger numerical errors during the computation of X_{LS} and is thus not advisable. The algorithm sketched in what follows derives X_{LS} by operating on A directly.

For any matrix A satisfying (A.8.11), there exist a unitary matrix $Q \in \mathbf{C}^{m\times m}$ and nonsingular upper triangular matrix $R \in \mathbf{C}^{n\times n}$ such that

$$A = Q \begin{bmatrix} R \\ 0 \end{bmatrix} \triangleq [\, \underbrace{Q_1}_{n} \quad \underbrace{Q_2}_{m-n} \,] \begin{bmatrix} R \\ 0 \end{bmatrix} \tag{A.8.18}$$

The previous factorization of A is called the *QR decomposition* (QRD). Inserting (A.8.18) into (A.8.14), we obtain

$$X_{LS} = R^{-1}Q_1^*B$$

Hence, once the QRD of A has been performed, X_{LS} can be obtained conveniently, as the solution of a triangular system of linear equations:

$$RX_{LS} = Q_1^*B \tag{A.8.19}$$

We note that the computation of the QRD is faster than that of the SVD (see, for example, [Stewart 1973; Golub and Van Loan 1989]).

The previous definition and derivation of X_{LS} make it clear that the LS approach derives an approximate solution to (A.8.12) by implicitly assuming that only the right-hand-side matrix, B, is perturbed. In applications, quite frequently *both A and B* are perturbed versions of some nominal (and unknown) matrices. In such cases, we may think of determining an approximate solution to (A.8.12) by explicitly recognizing the fact that neither A nor B is perturbation free. An approach based on this idea is described next (see, for example, [Van Huffel and Vandewalle 1991]).

Definition D17: The ***total least squares*** (TLS) approximate solution to (A.8.12) is defined as follows: First, derive the minimal perturbations Δ_A and Δ_B that make the system consistent—that is,

$$\min \ \|[\Delta_A \ \Delta_B]\|^2 \quad \text{subject to} \quad (A+\Delta_A)X = B + \Delta_B \tag{A.8.20}$$

Then, obtain X_{TLS} by solving the system in (A.8.20) corresponding to the optimal perturbations $(\Delta_A, \ \Delta_B)$.

A simple way to derive a more explicit formula for calculating the X_{TLS} is as follows:

Result R33: Let

$$[A \ B] = [\ \underbrace{\tilde{U}_1}_{n} \quad \underbrace{\tilde{U}_2}_{m-n}\ ] \begin{bmatrix} \tilde{\Sigma}_1 & 0 \\ 0 & \tilde{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \tilde{V}_1^* \\ \tilde{V}_2^* \end{bmatrix} \begin{matrix} \}\, n \\ \}\, p \end{matrix} \tag{A.8.21}$$

denote the SVD of the matrix $[A \ B]$. Furthermore, partition $\tilde{V}_2^*$ as

$$\tilde{V}_2^* = [\ \underbrace{\tilde{V}_{21}^*}_{n} \quad \underbrace{\tilde{V}_{22}^*}_{p}\ ] \tag{A.8.22}$$

Then

$$X_{TLS} = -\tilde{V}_{21} \ \tilde{V}_{22}^{-1} \tag{A.8.23}$$

if $\tilde{V}_{22}^{-1}$ exists.

Proof: The optimization problem with constraints in (A.8.20) can be restated in the following way: Find the minimal perturbation $[\Delta_A \ \Delta_B]$ and the corresponding matrix X such that

$$\{\ [A \ B] + [\Delta_A \ \Delta_B]\ \} \begin{bmatrix} -X \\ I \end{bmatrix} = 0 \tag{A.8.24}$$

Because $\text{rank}\begin{bmatrix} -X \\ I \end{bmatrix} = p$, $[\Delta_A \ \ \Delta_B]$ should be such that $\dim \mathcal{N}(\ [A \ \ B] + [\Delta_A \ \ \Delta_B]\ ) \geq p$ or, equivalently,

$$\text{rank}(\ [A \ B] + [\Delta_A \ \Delta_B]\ ) \leq n \tag{A.8.25}$$

According to R18, the minimal-perturbation matrix $[\Delta_A \ \Delta_B]$ that achieves (A.8.25) is given by

$$[\Delta_A \ \Delta_B] = -\tilde{U}_2 \tilde{\Sigma}_2 \tilde{V}_2^* \tag{A.8.26}$$

Inserting (A.8.26) along with (A.8.21) into (A.8.24), we obtain the following matrix equation in X:

$$\tilde{U}_1 \tilde{\Sigma}_1 \tilde{V}_1^* \begin{bmatrix} -X \\ I \end{bmatrix} = 0$$

Equivalently we have

$$\tilde{V}_1^* \begin{bmatrix} -X \\ I \end{bmatrix} = 0 \tag{A.8.27}$$

Equation (A.8.27) implies that X must satisfy

$$\begin{bmatrix} -X \\ I \end{bmatrix} = \tilde{V}_2 Q = \begin{bmatrix} \tilde{V}_{21} \\ \tilde{V}_{22} \end{bmatrix} Q \tag{A.8.28}$$

for some nonsingular normalizing matrix Q. The expression (A.8.23) for X_{TLS} is readily obtained from (A.8.28). ■

The TLS solution in (A.8.23) is unique if and only if the singular values $\{\tilde{\sigma}_k\}$ of the matrix $[A\ B]$ are such that $\tilde{\sigma}_n > \tilde{\sigma}_{n+1}$ (this follows from R18). When $\tilde{V}_{22}$ is singular, the TLS solution does not exist; see [Van Huffel and Vandewalle 1991].

The computation of the X_{TLS} requires the SVD of the $m \times (n+p)$ matrix $[A\ B]$. The solution X_{TLS} can be rewritten in a slightly different form. Let $\tilde{V}_{11}$, $\tilde{V}_{12}$ be defined via the following partition of $\tilde{V}_1^*$:

$$\tilde{V}_1^* = [\, \underbrace{\tilde{V}_{11}}_{n} \quad \underbrace{\tilde{V}_{12}}_{p} \,]$$

The orthogonality condition $\tilde{V}_1^* \tilde{V}_2 = 0$ can be rewritten as

$$\tilde{V}_{11} \tilde{V}_{21} + \tilde{V}_{12} \tilde{V}_{22} = 0$$

which yields

$$X_{TLS} = -\tilde{V}_{21} \tilde{V}_{22}^{-1} = \tilde{V}_{11}^{-1} \tilde{V}_{12} \tag{A.8.29}$$

Because p is usually (much) smaller than n, the formula (A.8.23) for X_{TLS} can often be more efficient computationally than is (A.8.29). (For example, in the common case of $p = 1$, (A.8.23) does not require a matrix inversion, whereas (A.8.29) requires the calculation of an $n \times n$ matrix inverse.)

A.9 QUADRATIC MINIMIZATION

Several problems in this text require the solution to *quadratic minimization problems*. In this section, we make use of matrix-analysis techniques to derive two results: one on unconstrained minimization, the other on constrained minimization.

Result R34: Let A be an $(n \times n)$ Hermitian positive definite matrix, let X and B be $(n \times m)$ matrices, and let C be an $m \times m$ Hermitian matrix. Then the unique solution to the minimization problem

$$\min_X F(X), \quad F(X) = X^*AX + X^*B + B^*X + C \tag{A.9.1}$$

is given by

$$X_0 = -A^{-1}B, \qquad F(X_0) = C - B^*A^{-1}B \tag{A.9.2}$$

Here, the matrix minimization means $F(X_0) \leq F(X)$ for every $X \neq X_0$; that is, $F(X) - F(X_0)$ is a positive semidefinite matrix.

Proof: Let $X = X_0 + \Delta$, where Δ is an arbitrary $(n \times m)$ complex matrix. Then

$$\begin{aligned} F(X) &= (-A^{-1}B + \Delta)^*A(-A^{-1}B + \Delta) + (-A^{-1}B + \Delta)^*B \\ &\quad + B^*(-A^{-1}B + \Delta) + C \\ &= \Delta^*A\Delta + F(X_0) \end{aligned} \tag{A.9.3}$$

Now, A is positive definite, so $\Delta^*A\Delta \geq 0$ for all nonzero Δ; thus, the minimum value of $F(X)$ is $F(X_0)$, and the result is proven. ■

We next present a result on linearly constrained quadratic minimization.

Result R35: Let A be an $(n \times n)$ Hermitian positive definite matrix, and let $X \in \mathbf{C}^{n \times m}$, $B \in \mathbf{C}^{n \times k}$, and $C \in \mathbf{C}^{m \times k}$. Assume that B has full column rank equal to k (hence $n \geq k$). Then the unique solution to the minimization problem

$$\min_X X^*AX \quad \text{subject to} \quad X^*B = C \tag{A.9.4}$$

is given by

$$X_0 = A^{-1}B(B^*A^{-1}B)^{-1}C^* \tag{A.9.5}$$

Proof: First, note that $(B^*A^{-1}B)^{-1}$ exists and that $X_0^*B = C$. Let $X = X_0 + \Delta$, where $\Delta \in \mathbf{C}^{n \times m}$ satisfies $\Delta^*B = 0$ (so that X also satisfies the constraint $X^*B = C$). Then

$$X^*AX = X_0^*AX_0 + X_0^*A\Delta + \Delta^*AX_0 + \Delta^*A\Delta \tag{A.9.6}$$

where the two middle terms are equal to zero:

$$\Delta^* A X_0 = \Delta^* B (B^* A^{-1} B)^{-1} C^* = 0$$

Hence,

$$X^* A X - X_0^* A X_0 = \Delta^* A \Delta \geq 0 \tag{A.9.7}$$

because A is positive definite. It follows from (A.9.7) that the minimizing X matrix is given by X_0. ■

A common special case of Result R35 is $m = k = 1$ (so X and B are both vectors) and $C = 1$. Then

$$X_0 = \frac{A^{-1} B}{B^* A^{-1} B}$$