Appendix A

USEFUL MATRIX RESULTS

Thomas Kailath

Ali H. Sayed

We collect in this appendix several matrix facts and formulas, where we assume that inverses exist as needed.

A.1 SOME MATRIX IDENTITIES

(i) Block Gaussian Elimination and Schur Complements

Consider a block matrix

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

that we wish to triangularize by a (block) Gaussian elimination procedure. For this, note that

$$\begin{bmatrix} I & 0 \\ X & I \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A & B \\ XA+C & XB+D \end{bmatrix},$$

so that choosing $X = -CA^{-1}$ gives

$$\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A & B \\ 0 & \Delta_A \end{bmatrix},$$

where

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

is called the *Schur complement of A in M*. Similarly we can find that

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ C & \Delta_A \end{bmatrix}.$$

Thus we also can obtain

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

and

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix} = \begin{bmatrix} \Delta_D & B \\ 0 & D \end{bmatrix},$$

where $\Delta_D = A - BD^{-1}C$ is the *Schur complement* of D in M.

(ii) Determinants

Using the product rule for determinants, the results in (i) give

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det A \det (D - CA^{-1}B) = \det A \det \Delta_A \tag{A.1.1}$$

$$= \det D \det (A - BD^{-1}C) = \det D \det \Delta_D. \tag{A.1.2}$$

(iii) **Block Triangular Factorizations**

The results in (i) can be combined to block diagonalize M:

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

and

$$\begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix}=\begin{bmatrix} \Delta_D & 0 \\ 0 & D \end{bmatrix}.$$

Then by using the easily verified formula

$$\begin{bmatrix} I & 0 \\ P & I \end{bmatrix}^{-1}=\begin{bmatrix} I & 0 \\ -P & I \end{bmatrix},$$

we can obtain the direct factorization formulas

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

(iv) **Recursive Triangularization and LDU Decomposition**

An alternative way of writing the above formulas is

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix}A^{-1}\begin{bmatrix} A & B \end{bmatrix}+\begin{bmatrix} 0 & 0 \\ 0 & \Delta_A \end{bmatrix} \tag{A.1.3}$$

$$= \begin{bmatrix} B \\ D \end{bmatrix}D^{-1}\begin{bmatrix} C & D \end{bmatrix}+\begin{bmatrix} \Delta_D & 0 \\ 0 & 0 \end{bmatrix}, \tag{A.1.4}$$

which also serves to *define* the Schur complements Δ_A and Δ_D. The above formulas can be used recursively to obtain, respectively, the block lower upper and block upper lower triangular factorizations of the matrix on the left-hand side.

In particular, by choosing A to be scalar and proceeding recursively we can obtain the important LDU decomposition of a *strongly* regular matrix, i.e., one whose leading minors are all nonzero. To demonstrate this so-called Schur reduction procedure, let R be an $n \times n$ strongly regular matrix whose individual entries we denote by r_{ij}. Let l_0 and u_0 denote the first column and the first row of R, respectively. In view of (A.1.3), we see that if we subtract from R the rank 1 matrix $l_0 r_{00}^{-1} u_0$, then we obtain a new matrix whose first row and column are zero,

$$R - l_0 r_{00}^{-1} u_0 = \begin{bmatrix} 0 & 0 \\ 0 & R_1 \end{bmatrix}.$$

The matrix R_1 is the Schur complement of R with respect to its $(0,0)$ entry r_{00}. Now, let $\{r_{ij}^{(1)}, l_1, u_1\}$ denote the entries, the first column, and the first row of R_1, respectively, and repeat the above procedure. In general, we can write for the jth step

$$R_j - l_j \left[r_{00}^{(j)}\right]^{-1} u_j = \begin{bmatrix} 0 & 0 \\ 0 & R_{j+1} \end{bmatrix}.$$

We conclude that we can express R in terms of the successive $\{l_i, u_i, r_{00}^{(i)}\}$ as follows:

$$\begin{aligned} R &= l_0 r_{00}^{-1} u_0 + \begin{bmatrix} 0 \\ l_1 \end{bmatrix} \left[r_{00}^{(1)}\right]^{-1} \begin{bmatrix} 0 & u_1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ l_2 \end{bmatrix} \left[r_{00}^{(2)}\right]^{-1} \begin{bmatrix} 0 & 0 & u_2 \end{bmatrix} + \cdots \\ &\triangleq LD^{-1}U, \end{aligned}$$

where L is lower triangular, D^{-1} is diagonal, and U is upper triangular. The nonzero parts of the columns of L are the $\{l_i\}_{i=0}^{n-1}$, while the nonzero parts of the rows of U are the $\{u_i\}_{i=0}^{n-1}$. Likewise, the entries of D are the $\{r_{00}^{(i)}\}_{i=0}^{n-1}$. We can further normalize the diagonal entries of L and U and define $\bar{L} = LD^{-1}$ and $\bar{U} = D^{-1}U$. In this case, we obtain $R = \bar{L}D\bar{U}$ and the diagonal entries of $\bar{L}$ and $\bar{U}$ are unity.

It is also immediate to verify that the LDU factorization of a strongly regular matrix is unique. Indeed, assume there exist two decompositions of the form $R = L_1 D_1 U_1 = L_2 D_2 U_2$, where $\{L_1, L_2\}$ are lower triangular with unit diagonal, $\{D_1, D_2\}$ are diagonal, and $\{U_1, U_2\}$ are upper triangular with unit diagonal. Then it must hold that

$$L_2^{-1} L_1 = D_2 U_2 U_1^{-1} D_1^{-1}.$$

Now the left-hand matrix in the above equality is lower triangular, while the right-hand matrix is upper triangular. Hence, equality holds only if both matrices are diagonal. But since the diagonal entries of $L_2^{-1}L_1$ are unity, it follows that we must have

$$L_2^{-1} L_1 = D_2 U_2 U_1^{-1} D_1^{-1} = I, \quad \text{the identity matrix,}$$

from which we conclude that $L_1 = L_2$, $D_1 = D_2$, and $U_1 = U_2$.

(v) Inverses of Block Matrices

When the block matrix is invertible, we can use the factorizations in (iii) to write

$$\begin{aligned} \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} &= \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ 0 & \Delta_A^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \\ &= \begin{bmatrix} A^{-1} + A^{-1}B\Delta_A^{-1}CA^{-1} & -A^{-1}B\Delta_A^{-1} \\ -\Delta_A^{-1}CA^{-1} & \Delta_A^{-1} \end{bmatrix}. \end{aligned} \quad \text{(A.1.5)}$$

Alternatively, we can write

$$\begin{aligned} \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} &= \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix} \begin{bmatrix} \Delta_D^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix} \begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix} \\ &= \begin{bmatrix} \Delta_D^{-1} & -\Delta_D^{-1}BD^{-1} \\ -D^{-1}C\Delta_D^{-1} & D^{-1} + D^{-1}C\Delta_D^{-1}BD^{-1} \end{bmatrix}. \end{aligned} \quad \text{(A.1.6)}$$

By equating the (1,1) and (2,2) elements in the right-hand sides of (A.1.5) and (A.1.6), we note that

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

(vi) More Inverse Formulas

Another useful set of formulas can be obtained from the formulas in (v) (and a little algebra):

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

and similarly

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

We also have formulas analogous to (A.1.3) and (A.1.4):

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -A^{-1}B \\ I \end{bmatrix}\Delta_A^{-1}\begin{bmatrix} -CA^{-1} & I \end{bmatrix}, \qquad \text{(A.1.7)}$$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 0 \\ 0 & D^{-1} \end{bmatrix} + \begin{bmatrix} I \\ -D^{-1}C \end{bmatrix}\Delta_D^{-1}\begin{bmatrix} I & -BD^{-1} \end{bmatrix}. \qquad \text{(A.1.8)}$$

(vii) The Matrix Inversion Lemma

For convenience of recall, replacing C by $-D$ and D by C^{-1}, we can rewrite the above formula for Δ_D^{-1} as

$$(A+BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1}+DA^{-1}B)^{-1}DA^{-1},$$

which is often called the modified matrices formula or the matrix inversion lemma.

(viii) Hermitian Matrices

The Hermitian conjugate A^* of a matrix A is the complex conjugate of its transpose. Hermitian matrices are (necessarily square) matrices obeying $A^* = A$. Such matrices have real eigenvalues, say, $\{\lambda_i\}$, and a full set of orthonormal eigenvectors, say, $\{p_i\}$. The so-called spectral decomposition of a Hermitian matrix is the representation

$$A = P\Lambda P^* = \sum_{i=1}^{n} \lambda_i p_i p_i^*,$$

where

$$\Lambda = \text{diag}\{\lambda_1,\ldots,\lambda_n\}, \quad P = \begin{bmatrix} p_1 & \ldots & p_n \end{bmatrix}, \quad Ap_i = \lambda_i p_i, \; i = 1,\ldots,n.$$

For strongly regular Hermitian matrices, the LDU decomposition takes the form

$$A = LDL^*, \quad L = \text{lower triangular with unit diagonal.}$$

The proof is instructive. If $A = LDU$, then $A^* = U^*D^*L^* = U^*DL^*$, since D is real valued. But by uniqueness of triangular factorization, we must have $U = L^*$.

(ix) Inertia Properties

Since the eigenvalues of a Hermitian matrix $A = A^*$ are real, we can define the inertia of A as the triple $In\{A\} = \{n_+, n_-, n_0\}$, where n_+ is the number of positive (> 0) eigenvalues of A, n_- is the number of negative (< 0) eigenvalues of A, and n_0 is the number of zero eigenvalues of A. Note that $n_+ + n_-$ = the rank of A, while n_0 is often called the *nullity of* A. The *signature* of A is the pair $\{n_+, n_-\}$. We shall define

$$\begin{aligned} S_A &\triangleq \text{the signature matrix of } A \\ &= \text{a diagonal matrix with } n_+ \text{ ones } (+1) \text{ and} \\ &\quad\; n_- \text{ minus ones } (-1) \text{ on the diagonal.} \end{aligned}$$

It is not necessary to compute the eigenvalues of A to determine its inertia or its signature matrix. The following property shows that it suffices to compute the LDL^* decomposition of A.

Lemma A.1.1 (Sylvester's Law of Inertia). *For any nonsingular matrix B, $In\{A\} = In\{BAB^*\}$.*

◇

The matrices A and BAB^* are said to be *congruent* to each other, so Sylvester's law states that congruence preserves inertia. The following useful result follows easily from the above and the factorizations in (iii).

Lemma A.1.2 (Inertia of Block Hermitian Matrices). *Let*

$$M = \begin{bmatrix} A & C^* \\ C & D \end{bmatrix}, \quad A = A^*, \quad D = D^*.$$

If A is nonsingular, then

$$In\{M\} = In\{A\} + In\{\Delta_A\}, \tag{A.1.9}$$

where $\Delta_A \triangleq D - CA^{-1}C^*$, the Schur complement of A in M. If D is nonsingular, then

$$In\{M\} = In\{D\} + In\{\Delta_D\}, \tag{A.1.10}$$

where $\Delta_D \triangleq A - C^*D^{-1}C$, the Schur complement of D in M.

◇

(x) Positive-Definite Matrices

An $n \times n$ Hermitian matrix A is positive semidefinite (p.s.d.) or nonnegative definite (n.n.d.), written $A \geq 0$, if it satisfies

$$x^*Ax \geq 0 \quad \text{for all } x \in \mathbb{C}^n.$$

It is strictly positive definite (p.d.), written $A > 0$, if $x^*Ax > 0$ except when $x = 0$.

Among the several characterizations of $A \geq 0$, we note the nonnegativity of all its eigenvalues and the fact that *all* minors are nonnegative. For strict positive definiteness it is necessary and sufficient that the leading minors be positive. An often more computationally useful characterization is that n.n.d. matrices can be factored as LDL^* or UDU^*, where all entries of the diagonal matrix D are nonnegative.

From the results of (ix), we note that a Hermitian block matrix

$$M = \begin{bmatrix} A & C^* \\ C & D \end{bmatrix}$$

is positive definite if and only if either $A > 0$ and $\Delta_A > 0$ or $D > 0$ and $\Delta_D > 0$.

A.2 THE GRAM–SCHMIDT PROCEDURE AND THE QR DECOMPOSITION

A fundamental step in many algorithms is that of replacing a collection of linearly independent vectors by a collection of orthonormal vectors that span the same column space.

Given n independent columns $\{a_i \in \mathbb{C}^N\}$, the so-called Gram–Schmidt procedure finds n orthonormal column vectors $\{q_i \in \mathbb{C}^N\}$ such that, for any $0 \leq j < n$, the column span of $\{q_0, \ldots, q_j\}$ coincides with the column span of $\{a_0, \ldots, a_j\}$. These vectors are determined recursively as follows. Start with $q_0 = a_0/\sqrt{a_0^* a_0}$. Now assume for step i that we have already found the orthonormal vectors $\{q_0, q_1, \ldots, q_{i-1}\}$. We project a_i onto the space spanned by these vectors and determine the residual vector r_i,

$$r_i = a_i - (q_0^* a_i)q_0 - (q_1^* a_i)q_1 - \cdots - (q_{i-1}^* a_i)q_{i-1}.$$

We further scale r_i to have unit norm and take the result to be q_i:

$$q_i = r_i/\sqrt{r_i^* r_i}.$$

(The fact that the vectors $\{a_k\}$ are linearly independent guarantees a nonzero r_i.) It follows from this construction that each a_i can be expressed as a linear combination of the resulting $\{q_0, q_1, \ldots, q_i\}$, viz.,

$$a_i = (q_0^* a_i)q_0 + (q_1^* a_i)q_1 + \cdots + (q_{i-1}^* a_i)q_{i-1} + \sqrt{r_i^* r_i}\, q_i.$$

If we now introduce the $N \times n$ matrices

$$A = \text{col}\{a_0, a_1, \ldots, a_{n-1}\}, \quad \bar{Q} = \text{col}\{q_0, q_1, \ldots, q_{n-1}\},$$

we conclude that the above construction leads to the so-called reduced QR factorization of A,

$$A = \bar{Q}\bar{R},$$

where $\bar{R}$ is $n \times n$ upper triangular,

$$\bar{R} = \begin{bmatrix} \sqrt{r_0^* r_0} & q_0^* a_1 & q_0^* a_2 & \cdots & q_0^* a_{n-1} \\ & \sqrt{r_1^* r_1} & q_1^* a_2 & \cdots & q_1^* a_{n-1} \\ & & \ddots & & \vdots \\ & & & \ddots & \vdots \\ & & & & \sqrt{r_{n-1}^* r_{n-1}} \end{bmatrix}.$$

A *full* QR factorization of A can be obtained by appending an additional $N-n$ orthonormal columns to $\bar{Q}$ so that it becomes a unitary $N \times N$ matrix. Likewise, we append rows of zeros to $\bar{R}$ so that it becomes an $N \times n$ matrix:

$$A = QR \equiv \begin{bmatrix} \bar{Q} & q_n & \dots & q_{N-1} \end{bmatrix} \begin{bmatrix} \bar{R} \\ 0 \end{bmatrix}.$$

The orthogonalization procedure so described is not reliable numerically due to the accumulation of round-off errors in finite precision arithmetic. A so-called modified Gram–Schmidt procedure has better numerical properties. It operates as follows:

1. We again start with $q_0 = a_0/\sqrt{a_0^* a_0}$ but now project the remaining column vectors $\{a_1, a_2, \dots, a_{n-1}\}$ onto q_0. The corresponding residuals are denoted by $a_j^{(1)} = a_j - (q_0^* a_j) q_0$. This step therefore replaces all the original vectors $\{a_0, \dots, a_{n-1}\}$ by the new vectors $\{q_0, a_1^{(1)}, a_2^{(1)}, \dots, a_{n-1}^{(1)}\}$.

2. We then take $q_1 = a_1^{(1)}/\sqrt{a_1^{*(1)} a_1^{(1)}}$ and project the remaining column vectors $\{a_2^{(1)}, \dots, a_{n-1}^{(1)}\}$ onto q_1. The corresponding residuals are denoted by $a_j^{(2)} = a_j^{(1)} - (q_1^* a_j^{(1)}) q_1$. This step replaces $\{q_0, a_1^{(1)}, a_2^{(1)}, \dots, a_{n-1}^{(1)}\}$ by $\{q_0, q_1, a_2^{(2)}, a_3^{(2)}, \dots, a_{n-1}^{(2)}\}$.

3. We now take $q_2 = a_2^{(2)}/\sqrt{a_2^{*(2)} a_2^{(2)}}$ and proceed as above.

For both variants of the Gram–Schmidt algorithm, we can verify that the computational cost involved is $O(2Nn^2)$ flops. We should further add that the QR factorization of a matrix can also be achieved by applying a sequence of numerically reliable rotations to the matrix (as explained, for example, in App. B).

A.3 MATRIX NORMS

The 2-induced norm of a matrix A, also known as the spectral norm of the matrix, is defined by

$$\|A\|_2 \triangleq \max_{\|x\| \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{\|z\|=1} \|Az\|,$$

where $\|x\|$ denotes the Euclidean norm of the vector x. It can be shown that $\|A\|_2$ is also equal to the maximum singular value of A. More specifically, the following two conclusions can be established. Let $\sigma_{\max}$ denote the largest singular value and let $\sigma_{\min}$ denote the smallest singular value. Then

$$\sigma_{\max} = \max_{\|x\|=1} \|Ax\| \quad \text{and} \quad \sigma_{\min} = \min_{\|x\|=1} \|Ax\|.$$

The Frobenius norm of a matrix $A = [a_{ij}]$ is defined by

$$\|A\|_{\mathrm{F}} \triangleq \sqrt{\sum_{i=0}^{M} \sum_{j=0}^{m} |a_{ij}|^2}. \tag{A.3.1}$$

In terms of the singular values of A, it is easy to verify that if A has rank p with nonzero singular values $\{\sigma_1, \dots, \sigma_p\}$, then

$$\|A\|_{\mathrm{F}} = \sqrt{\mathrm{trace}(A^* A)} = \sqrt{\sum_{i=1}^{p} \sigma_i^2}.$$

A.4 UNITARY AND J-UNITARY TRANSFORMATIONS

The following result plays a key role in the derivation of many array algorithms. One proof uses the SVDs of the involved matrices.

Lemma A.4.1 (Basis Rotation). *Given two $n \times m$ $(n \leq m)$ matrices A and B. Then $AA^* = BB^*$ if and only if there exists an $m \times m$ unitary matrix Θ $(\Theta\Theta^* = I = \Theta^*\Theta)$ such that $A = B\Theta$.*

Proof: One implication is immediate. If there exists a unitary matrix Θ such that $A = B\Theta$, then $AA^* = (B\Theta)(B\Theta)^* = B(\Theta\Theta^*)B^* = BB^*$. One proof for the converse implication follows by invoking the SVDs of A and B, say,

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

where U_A and U_B are $n \times n$ unitary matrices, V_A and V_B are $m \times m$ unitary 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^* (BB^*). Moreover, U_A (U_B) can be constructed from an orthonormal basis for the right eigenvectors of AA^* (BB^*). Hence, it follows from the identity $AA^* = BB^*$ that we have $\Sigma_A = \Sigma_B$ and $U_A = U_B$. Let $\Theta = V_B V_A^*$. We then get $\Theta\Theta^* = I$ and $B\Theta = A$.

$\diamond$

We can establish a similar result when the equality $AA^* = BB^*$ is replaced by $AJA^* = BJB^*$ for some signature matrix J. More specifically, we have the following statement.

Lemma A.4.2 (J-Unitary Transformations). *Let A and B be $n \times m$ matrices (with $n \leq m$), and let $J = (I_p \oplus -I_q)$ be a signature matrix with $p+q = m$. If $AJA^* = BJB^*$ is full rank, then there exists a J-unitary matrix Θ such that $A = B\Theta$.*

Proof: Since AJA^* is Hermitian and invertible,[9] we can factor it as $AJA^* = RSR^*$, where $R \in \mathbb{C}^{n\times n}$ is invertible and $S = (I_\alpha \oplus -I_\beta)$ is a signature matrix (with $\alpha + \beta = n$). We normalize A and B by defining $\bar{A} = R^{-1}A$ and $\bar{B} = R^{-1}B$. Then $\bar{A}J\bar{A}^* = \bar{B}J\bar{B}^* = S$.

Now consider the block triangular factorizations

$$\begin{bmatrix} S & \bar{A} \\ \bar{A}^* & J \end{bmatrix} = \begin{bmatrix} I & \\ \bar{A}^*S & I \end{bmatrix} \begin{bmatrix} S & \\ & J - \bar{A}^*S\bar{A} \end{bmatrix} \begin{bmatrix} I & \\ \bar{A}^*S & I \end{bmatrix}^*$$
$$= \begin{bmatrix} I & \bar{A}J \\ & I \end{bmatrix} \begin{bmatrix} \underbrace{S - \bar{A}J\bar{A}^*}_{=0} & \\ & J \end{bmatrix} \begin{bmatrix} I & \bar{A}J \\ & I \end{bmatrix}^*.$$

Using the fact that the central matrices must have the same inertia, we conclude that $In\{J - \bar{A}^*S\bar{A}\} = In\{J\} - In\{S\} = \{p-\alpha, q-\beta, n\}$. Similarly, we can show that $In\{J - \bar{B}^*S\bar{B}\} = \{p-\alpha, q-\beta, n\}$.

Define the signature matrix $J_1 \triangleq (I_{p-\alpha} \oplus -I_{q-\beta})$. The above inertia conditions then mean that we can factor $(J - \bar{A}^*S\bar{A})$ and $(J - \bar{B}^*S\bar{B})$ as

$$J - \bar{A}^*S\bar{A} = XJ_1X^*, \quad J - \bar{B}^*S\bar{B} = YJ_1Y^*, \qquad X, Y \in \mathbb{C}^{m\times m-n}.$$

[9] This argument was suggested by Professor T. Constantinescu.

Finally, introduce the square matrices

$$\Sigma_1 = \begin{bmatrix} \bar{A} \\ X^* \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} \bar{B} \\ Y^* \end{bmatrix}.$$

It is easy to verify that these matrices satisfy $\Sigma_1^*(S \oplus J_1)\Sigma_1 = J$ and $\Sigma_2^*(S \oplus J_1)\Sigma_2 = J$. Moreover, $\Sigma_1 J \Sigma_1^* = (S \oplus J_1)$ and $\Sigma_2 J \Sigma_2^* = (S \oplus J_2)$. These relations allow us to relate Σ_1 and Σ_2 as $\Sigma_1 = \Sigma_2[J\Sigma_2^*(S \oplus J_1)\Sigma_1]$. If we set $\Theta \triangleq [J\Sigma_2^*(S \oplus J_1)\Sigma_1]$, then it is immediate to check that Θ is J-unitary and, from the equality of the first block row of $\Sigma_1 = \Sigma_2\Theta$, that $\bar{A} = \bar{B}\Omega$. Hence, $A = B\Theta$.

◇

In the above statement, the arrays A and B are either square or fat ($n \leq m$). We can establish a similar result when $n \geq m$ instead. For this purpose, first note that if A is an $n \times m$ and full-rank matrix, with $n \geq m$, then its SVD takes the form

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

where Σ is $n \times n$ and invertible. The left inverse of A is defined by

$$A^\dagger \triangleq V \begin{bmatrix} \Sigma^{-1} & 0 \end{bmatrix} U^*,$$

and it satisfies $A^\dagger A = I_m$, the identify matrix of size m.

Lemma A.4.3 (J-Unitary Transformations). *Let A and B be $n \times m$ full-rank matrices (with $n \geq m$) and let $J = (I_p \oplus -I_q)$ be a signature matrix with $p + q = m$. The relation $AJA^* = BJB^*$ holds if and only if there exists a unique $m \times m$ J-unitary matrix Θ such that $A = B\Theta$.*

Proof: The "if" statement is immediate. For the converse, note that since A and B are assumed full rank, there exist left inverses $A^\dagger$ and $B^\dagger$ such that $A^\dagger A = I_m$ and $B^\dagger B = I_m$. Now define $\Theta = B^\dagger A$. We claim that Θ is J-unitary and maps B to A, as desired.

The proof that $\Theta J \Theta^* = J$ is immediate from the equality $AJA^* = BJB^*$. Just multiply it from the left by $B^\dagger$ and from the right by $(B^\dagger)^*$ and use $B^\dagger B = I_m$.

To prove that $B\Theta = A$, for the above choice of Θ, we start with $AJA^* = BJB^*$ again and insert the term $B^\dagger B$ into the right-hand side to get

$$AJA^* = B(B^\dagger B)JB^* = BB^\dagger(BJB^*) = BB^\dagger AJA^*.$$

Multiplying from the right by $(A^\dagger)^*$ and using $A^\dagger A = I_m$ we obtain $BB^\dagger(AJ) = AJ$. Since J is invertible and its inverse is J, we conclude by multiplying by J from the right that $B(B^\dagger A) = A$, which is the desired result. That is, Θ is J-unitary and maps B to A.

To show that Θ is unique, assume $\bar{\Theta}$ is another J-unitary matrix that maps B to A and write $B\Theta = B\bar{\Theta}$. Now multiply by $B^\dagger$ from the left to conclude that $\Theta = \bar{\Theta}$.

◇

A.5 Two Additional Results

Finally, we state two matrix results that are needed in Chs. 5 and 8. The first theorem is cited in Ch. 5. Its statement was provided by the authors of that chapter.

Theorem A.5.1 (Cauchy Interlace Theorem) (see [Par80]). *Let A, W, and Y be symmetric matrices and $A = W + Y$. Let the eigenvalues of the matrices be ordered as*

$$\lambda_1 \le \lambda_2 \le \cdots \lambda_n$$

or

$$\lambda_{-n} \le \cdots \le \lambda_{-2} \le \lambda_{-1}.$$

For any i, j satisfying $1 \le i + j - 1 \le n$, the following inequalities hold:

$$\lambda_i(W) + \lambda_j(Y) \le \lambda_{i+j-1}(A)$$

and

$$\lambda_{-(i+j-1)}(A) \le \lambda_{-i}(W) + \lambda_{-j}(Y).$$

◊

The next theorem is cited in Ch. 8. Its statement was provided by the authors of that chapter.

Theorem A.5.2 (Perron–Frobenius). *Let $A = (a_{i,j})$ be an $n \times n$ matrix with nonnegative entries and denote ρ its spectral radius, that is, the maximum modulus of its eigenvalues. Then*

1. *there exists an eigenvalue λ of A such that $\rho = \lambda$;*
2. *there exists an eigenvector v of A with nonnegative components corresponding to λ;*
3. *if the matrix is not reducible then $\lambda > 0$ and v has positive components; moreover, λ and v are unique. (A matrix A is said to be reducible if there exists a permutation of rows and columns that transforms A by similarity in the following way:*

$$PAP^T = \begin{bmatrix} A_{1,1} & A_{1,2} \\ O & A_{2,2} \end{bmatrix},$$

where the blocks $A_{1,1}$, $A_{2,2}$ are square matrices and P is a permutation matrix.)

◊