Philip J. Davis
Circulant Matrices
Chelsea Publishing New York
2nd edition 1994

2 INTRODUCTORY MATRIX MATERIAL

2.1 BLOCK OPERATIONS

It is very often convenient in both theoretical and computer work to partition a matrix into submatrices. This can be done in numerous ways as suggested by this example:

$$\left(\begin{array}{cc|cc} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ \hline 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{array}\right) \quad \left(\begin{array}{c|c|c|c} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ \hline 13 & 14 & 15 & 16 \end{array}\right).$$

Each submatrix or block can be labeled by subscripts, and we can display the original matrix with submatrices or blocks for its elements. The general form of a partitioned matrix therefore is

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1\ell} \\ \vdots & \vdots & & \vdots \\ A_{k1} & A_{k2} & \cdots & A_{k\ell} \end{pmatrix}. \tag{2.1.1}$$

Dotted lines, bars, commas are all used in an obvious way to indicate partitions. The size of the blocks must be such that they all fit together properly.

This means that the number of rows in each A_{ij} must be the same for each i and the number of columns must be the same for each j. The size of A_{ij} is therefore $m_i \times n_j$ for certain integers m_i and n_j. We indicate this by writing

$$A = \begin{array}{c} \begin{array}{cccc} \text{No. of columns} & & & \\ m_1 & m_2 & \cdots & m_\ell \end{array} \\ \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1\ell} \\ \vdots & \vdots & & \vdots \\ A_{k1} & A_{k2} & \cdots & A_{k\ell} \end{pmatrix} \end{array} \begin{array}{c} \text{No. of rows} \\ n_1 \\ \vdots \\ n_k \end{array} \tag{2.1.1'}$$

A square matrix A of order n is often partitioned symmetrically. Suppose that $n = n_1 + n_2 + \cdots + n_r$ with $n_i \geq 1$. Partition A as

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1r} \\ \vdots & & & \\ A_{r1} & A_{r2} & \cdots & A_{rr} \end{pmatrix} \tag{2.1.2}$$

where size $A_{ij} = n_i \times n_j$. The diagonal blocks A_{ii} are square matrices of order n_i.

Example.

$$\left(\begin{array}{cc|c|ccc} x & x & x & x & x & x \\ x & x & x & x & x & x \\ \hline x & x & x & x & x & x \\ \hline x & x & x & x & x & x \\ x & x & x & x & x & x \\ x & x & x & x & x & x \end{array}\right) \qquad \begin{array}{l} n = 6 \\ n_1 = 2 \\ n_2 = 1 \\ n_3 = 3 \end{array}$$

is a symmetric partition of a 6×6 matrix.

Square matrices are often built up, or compounded, of square blocks all of the same size.

Example.
$$\left(\begin{array}{ccc|ccc} x & x & x & x & x & x \\ x & x & x & x & x & x \\ x & x & x & x & x & x \\ \hline x & x & x & x & x & x \\ x & x & x & x & x & x \\ x & x & x & x & x & x \end{array}\right)$$

If a square matrix A of order nk is composed of n × n square submatrices all of order k, it is termed an (n, k) matrix. Thus the matrix depicted above is a (2, 3) matrix.

Subject to certain conformability conditions on the blocks, the operations of scalar product, transpose, conjugation, addition, and multiplication are carried out in the same way when expressed in block notation as when they are expressed in element notation. This means

$$(2.1.3)\qquad c\begin{pmatrix} A_{11} & \cdots & A_{1\ell} \\ \vdots & & \vdots \\ A_{k1} & \cdots & A_{k\ell} \end{pmatrix} = \begin{pmatrix} cA_{11} & \cdots & cA_{1\ell} \\ \vdots & & \vdots \\ cA_{k1} & \cdots & cA_{k\ell} \end{pmatrix},$$

$$(2.1.4)\qquad \begin{pmatrix} A_{11} & \cdots & A_{1\ell} \\ \vdots & & \vdots \\ A_{k1} & \cdots & A_{k\ell} \end{pmatrix}^{T} = \begin{pmatrix} A_{11}^{T} & & A_{k1}^{T} \\ \vdots & & \vdots \\ A_{1\ell}^{T} & \cdots & A_{k\ell}^{T} \end{pmatrix},$$

$$(2.1.5)\qquad \begin{pmatrix} A_{11} & \cdots & A_{1\ell} \\ \vdots & & \\ A_{k1} & \cdots & A_{k\ell} \end{pmatrix}^{*} = \begin{pmatrix} A_{11}^{*} & \cdots & A_{k1}^{*} \\ \vdots & & \vdots \\ A_{1\ell}^{*} & & A_{k\ell}^{*} \end{pmatrix}.$$

Here T designates the transpose and * the conjugate transpose.

$$(2.1.6)\qquad \begin{pmatrix} A_{11} & \cdots & A_{1\ell} \\ \vdots & & \vdots \\ A_{k1} & \cdots & A_{k\ell} \end{pmatrix} + \begin{pmatrix} B_{11} & \cdots & B_{1\ell} \\ \vdots & & \vdots \\ B_{k1} & \cdots & B_{k\ell} \end{pmatrix}$$
$$= \begin{pmatrix} A_{11}+B_{11} & \cdots & A_{1\ell}+B_{1\ell} \\ \vdots & & \vdots \\ A_{k1}+B_{k1} & \cdots & A_{k\ell}+B_{k\ell} \end{pmatrix},$$

$$(2.1.7)\qquad \begin{pmatrix} A_{11} & \cdots & A_{1\ell} \\ \vdots & & \\ A_{k1} & \cdots & A_{k\ell} \end{pmatrix}\begin{pmatrix} B_{11} & \cdots & B_{1n} \\ \vdots & & \vdots \\ B_{\ell 1} & \cdots & B_{\ell n} \end{pmatrix} = \begin{pmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & & \vdots \\ C_{k1} & \cdots & C_{kn} \end{pmatrix}$$

where $C_{ij} = \sum_{r=1}^{\ell} A_{ir}B_{rj}$.

In (2.1.6) the size of each A_{ij} must be the size of the corresponding B_{ij}.

In (2.1.7), designate the size of A_{ij} by $\alpha_i \times \beta_j$ and the size of B_{ij} by $\gamma_i \times \delta_j$. Then, if $\beta_r = \gamma_r$ for $1 \le r \le \ell$, the product $A_{ir}B_{rj}$ can be formed and produces an $\alpha_i \times \delta_j$ matrix, independently of r. The sum can then be found as indicated and the C_{ij} are $\alpha_i \times \delta_j$ matrices and together constitute a partition. Note that the rule for forming the blocks C_{ij} of the matrix product is the same as when A_{ij} and B_{ij} are single numbers.

Example. If A and B are n × n matrices and if

$$C = \begin{pmatrix} A & B \\ B & -A \end{pmatrix},$$

then

$$C^2 = \begin{pmatrix} A^2 + B^2, & AB - BA \\ BA - AB, & A^2 + B^2 \end{pmatrix}.$$

PROBLEMS

1. In the example just given what is C^2 if A and B commute?
2. In the example, compute C^3. What if A and B commute?
3. Let

$$M = \begin{pmatrix} B_1 & & 0 \\ & B_2 & \\ & & \ddots \\ 0 & & B_r \end{pmatrix}, \quad N = \begin{pmatrix} C_1 & & 0 \\ & C_2 & \\ & & \ddots \\ 0 & & C_s \end{pmatrix}$$

be two block diagonal matrices. When can the product MN be formed? What is the product?

4. "Hadamard matrices" of order 2^n are given recursively by means of the definition

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \qquad H_{2^{k+1}} = \begin{bmatrix} H_{2^k} & H_{2^k} \\ H_{2^k} & -H_{2^k} \end{bmatrix}.$$

Write out H_4 and H_8 explicitly. Compute $H_2H_2^T$, $H_4H_4^T$.

5. Let A, B, C, D all be n × n and let a, b, c, d be scalars. What is

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \quad \begin{pmatrix} aI & bI \\ cI & dI \end{pmatrix} ?$$

6. Let I_p be the identity of order p. Prove that

$$\det \begin{pmatrix} I_p & B \\ 0 & C \end{pmatrix} = \det C.$$

7. If A and C are square, prove that $\det \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} = (\det A)(\det C)$.
8. If A and C are square, prove that the eigenvalues of $\begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$ are those of A together with those of C.

2.2 DIRECT SUMS

For i = 1, 2, ..., k, let A_i be a square matrix of order n_i. The block diagonal square matrix

$$(2.2.1) \quad A = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & A_k \end{pmatrix} = \operatorname{diag}(A_1, A_2, \ldots, A_k)$$

of order $n_1 + n_2 + \cdots + n_k$ is called the direct sum of the A_k and is designated by

$$(2.2.2) \quad A = A_1 \oplus A_2 \oplus \cdots \oplus A_k = \bigoplus_{i=1}^{k} A_i.$$

The following identities are easily established.

(1) $(A \oplus B) \oplus C = A \oplus (B \oplus C)$.

(2) $(A + B) \oplus (C + D) = (A \oplus C) + (B \oplus D)$.

(3) $(A \oplus B)(C \oplus D) = AC \oplus BD$.

(4) $(A \oplus B)^T = A^T \oplus B^T$.

(5) $(A \oplus B)^* = A^* \oplus B^*$.

(6) $(A \oplus B)^{-1} = A^{-1} \oplus B^{-1}$, assuming that the indicated inverses exist.

(7) $\det(A \oplus B) = (\det A)(\det B)$.

(8) $\operatorname{tr}(A \oplus B) = \operatorname{tr} A + \operatorname{tr} B$.

(9) If $p_A(\lambda)$ designates the characteristic polynomial of A, then $p_{A \oplus B}(\lambda) = (p_A(\lambda))(p_B(\lambda))$.

(10) Hence $\lambda(A \oplus B) = \{\lambda A, \lambda B\}$. ($\lambda A$ designates the set of eigenvalues of A.)

PROBLEMS

1. Let $A = A_1 \oplus A_2 \oplus \cdots \oplus A_k$. Prove that $\det A = \prod_{i=1}^{k} \det A_i$ and that for integer p, $A^p = A_1^p \oplus A_2^p \oplus \cdots \oplus A_k^p$.

2. Give a linear algebra interpretation of the direct sum along the following lines. Let V be a finite-dimensional vector space and let L and M be subspaces. Write $V = L \oplus M$ if and only if every vector $x \in V$ can be written uniquely in the form $x = y + z$ with $y \in L$, $z \in M$. Show that $V = L \oplus M$ if and only if

   (a) $\dim V = \dim L + \dim M$, $L \cap M = \{0\}$.
   (b) if $\{x_1, \ldots, x_\ell\}$ and $\{y_1, \ldots, y_m\}$ are bases for L and M, then $\{x_1, \ldots, x_\ell, y_1, \ldots, y_m\}$ is a basis for V.

3. The fundamental theorem of rank-canonical form for square matrices tells us that if A is a $n \times n$ matrix of rank r, then there exist nonsingular matrices P, Q such that $PAQ = I_r \oplus 0_{n-r}$. Verify this formulation.

2.3 KRONECKER PRODUCT

Let A and B be $m \times n$ and $p \times q$ respectively. Then the Kronecker product (or tensor, or direct product of A and B) is that $mp \times nq$ matrix defined by

$$A \otimes B = \begin{pmatrix} a_{11}B, & a_{12}B, & \ldots, & a_{1n}B \\ \vdots & & & \\ a_{m1}B, & a_{m2}B, & \ldots, & a_{mn}B \end{pmatrix}. \tag{2.3.1}$$

Important properties of the Kronecker product are as follows (indicated operations are assumed to be defined):

(1) $(\alpha A) \otimes B = A \otimes (\alpha B) = \alpha(A \otimes B)$; α scalar.
(2) $(A + B) \otimes C = (A \otimes C) + (B \otimes C)$.
(3) $A \otimes (B + C) = (A \otimes B) + (A \otimes C)$.
(4) $A \otimes (B \otimes C) = (A \otimes B) \otimes C$.

(5) $(A \otimes B)(C \otimes D) = (AC) \otimes BD$.
(6) $\overline{A \otimes B} = \overline{A} \otimes \overline{B}$.
(7) $(A \otimes B)^T = A^T \otimes B^T$; $(A \otimes B)^* = A^* \otimes B^*$.
(8) $r(A \otimes B) = r(A)r(B)$.

We now assume that A and B are square and of orders m and n. Then

(9) $\operatorname{tr}(A \otimes B) = (\operatorname{tr}(A))(\operatorname{tr}(B))$.
(10) If A and B are nonsingular, so is $A \otimes B$ and $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.
(11) $\det(A \otimes B) = (\det A)^n (\det B)^m$.
(12) There exists a permutation matrix P (see Section 2.4) depending only on m, n, such that $B \otimes A = P^*(A \otimes B)P$.
(13) Let $\phi(x, y)$ designate the polynomial

$$\phi(x, y) = \sum_{j,k=0}^{p} a_{jk} x^j y^k.$$

Let $\phi(A; B)$ designate the $mn \times mn$ matrix

$$\sum_{j,k=0}^{p} a_{j,k} A^j \otimes B^k.$$

Then the eigenvalues of $\phi(A; B)$ are $\phi(\lambda_r, \mu_s)$, $r = 1, 2, \ldots, m$, $s = 1, 2, \ldots, n$ where λ_r and μ_s are the eigenvalues of A and B respectively. In particular, the eigenvalues of $A \otimes B$ are $\lambda_r \mu_s$, $r = 1, 2, \ldots m$; $s = 1, 2, \ldots, n$.

PROBLEMS

1. Show that $I_m \otimes I_n = I_{mn}$.
2. Describe the matrices $I \otimes A$, $A \otimes I$.
3. If A is $m \times m$ and B is $n \times n$, then $A \otimes B = (A \otimes I_n)(I_m \otimes B) = (I_m \otimes B)(A \otimes I_n)$.

4. If A and B are upper (or lower) triangular, then so is $A \otimes B$.
5. If $A \otimes B \neq 0$ is diagonal, so are A and B.
6. Let A and B have orders m, n respectively. Show that the matrix $(I_m \otimes B) + (A \otimes I_n)$ has the eigenvalues $\lambda_r + \mu_s$, $r = 1, 2, \ldots, m$, $s = 1, 2, \ldots, n$, where λ_r and μ_s are the eigenvalues of A and B. This matrix is often called the Kronecker sum of A and B.
7. Let A and B be of orders m and n. If A and B both are (1) normal, (2) Hermitian, (3) positive definite, (4) positive semidefinite, and (5) unitary, then $A \otimes B$ has the corresponding property. See Section 2.9.
8. Kronecker powers: Let $A^{[2]} = A \otimes A$ and, in general, $A^{[k+1]} = A \otimes A^{[k]}$. Prove that $A^{[k+\ell]} = A^{[k]} \otimes A^{[\ell]}$.
9. Prove that $(AB)^{[k]} = A^{[k]}B^{[k]}$.
10. Let $Ax = \lambda x$ and $By = \mu y$, $x = (x_1, \ldots, x_n)^T$. Define Z by $Z^T = [x_1y^T, x_2y^T, \ldots, x_my^T]$. Prove that $(A \otimes B)Z = \lambda\mu Z$.

2.4 PERMUTATION MATRICES

By a permutation σ of the set $N = \{1, 2, \ldots, n\}$ is meant a one-to-one mapping of N onto itself. Including the identity permutation there are n! distinct permutations of N. One can indicate a typical permutation by

$$\begin{aligned} \sigma(1) &= i_1 \\ \sigma(2) &= i_2 \\ &\vdots \\ \sigma(n) &= i_n \end{aligned} \tag{2.4.1}$$

which is often written as

$$\sigma: \begin{pmatrix} 1 & 2 & \ldots & n \\ i_1 & i_2 & \ldots & i_n \end{pmatrix}. \tag{2.4.1'}$$

The inverse permutation is designated by σ^{-1}. Thus $\sigma^{-1}(i_k) = k$.

Let E_j designate the unit (row) vector of n components which has a 1 in the jth position and 0's elsewhere:

$$E_j = (0, \ldots, 0, 1, 0, \ldots, 0). \tag{2.4.2}$$

By a permutation matrix of order n is meant a matrix of the form

$$P = P_\sigma = \begin{pmatrix} E_{i_1} \\ E_{i_2} \\ \vdots \\ E_{i_n} \end{pmatrix}. \tag{2.4.3}$$

One has

$$P = (a_{ij}) \quad \text{where} \quad \begin{aligned} a_{i,\sigma(i)} &= 1, \ i = 1,2,\ldots \\ a_{i,j} &= 0, \ \text{otherwise.} \end{aligned} \tag{2.4.4}$$

The ith row of P has a 1 in the $\sigma(i)$th column and 0's elsewhere. The jth column of P has a 1 in the $\sigma^{-1}(j)$th row and 0's elsewhere. Thus each row and each column of P has precisely one 1 in it.

Example

$$\sigma: \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix}, \qquad P_\sigma = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

It is easily seen that

$$(2.4.5)\qquad P_\sigma \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_{\sigma(1)} \\ x_{\sigma(2)} \\ \vdots \\ x_{\sigma(n)} \end{pmatrix}.$$

Hence if $A = (a_{ij})$ is an $n \times r$ matrix,

$$(2.4.6)\qquad \boxed{P_\sigma A = (a_{\sigma(i),j}),}$$

that is, $P_\sigma A$ is A with its rows permuted by σ. Moreover,

$$(2.4.7)\qquad (x_1, x_2, \ldots, x_n)P_\sigma = (x_{\sigma^{-1}(1)}, x_{\sigma^{-1}(2)}, \ldots, x_{\sigma^{-1}(n)}),$$

so that if $A = (a_{ij})$ is $r \times n$,

$$(2.4.8)\qquad \boxed{AP_\sigma = (a_{i,\sigma^{-1}(j)}).}$$

That is, AP_σ is A with its columns permuted by σ^{-1}.

Note also that

$$(2.4.9)\qquad \boxed{P_\sigma P_\tau = P_{\sigma\tau},}$$

where the product of the permutations σ, τ is applied from left to right. Furthermore,

$$(2.4.10)\qquad (P_\sigma)^* = P_{\sigma^{-1}};$$

hence

$$(2.4.11)\qquad (P_\sigma)^* P_\sigma = P_{\sigma^{-1}} P_\sigma = P_I = I.$$

Therefore

$$(2.4.12)\qquad \boxed{(P_\sigma)^* = P_{\sigma^{-1}} = (P_\sigma)^{-1}.}$$

The permutation matrices are thus unitary, forming a subgroup of the unitary group.

From (2.4.6), (2.4.8) and (2.4.12) it follows that if A is $n \times n$

$$(2.4.13)\qquad \boxed{P_\sigma A P_\sigma^* = (a_{\sigma(i),\sigma(j)}),}$$

so that the similarity transformation $P_\sigma A P_\sigma^*$ causes a consistent renumbering of the rows and columns of A by the permutation σ.

Among the permutation matrices, the matrix

$$(2.4.14)\qquad \boxed{\pi = \begin{pmatrix} 0 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & \ldots & & \ldots & 0 \end{pmatrix}}$$

plays a fundamental role in the theory of circulants. This corresponds to the forward shift permutation $\sigma(1) = 2$, $\sigma(2) = 3$, ..., $\sigma(n-1) = n$, $\sigma(n) = 1$, that is, to the cycle $\sigma = (1, 2, 3, \ldots, n)$ generating the cyclic group of order n (π is for "push"). One has

$$(2.4.15)\qquad \pi^2 = \begin{pmatrix} 0 & 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 1 & 0 & 0 & \ldots & 0 \end{pmatrix}$$

corresponding to σ^2 for which $\sigma^2(1) = 3$, $\sigma^2(2) = 4$, ..., $\sigma^2(n) = 2$. Similarly for π^k and σ^k. The matrix π^n corresponds to $\sigma^n = I$, so that

$$(2.4.16)\qquad \pi^n = I.$$

Note also that

$$(2.4.17)\qquad \pi^T = \pi^* = \pi^{-1} = \pi^{n-1}.$$

A particular instance of (2.4.13) is

$$(2.4.18)\qquad \boxed{\pi A \pi^T = (a_{i+1,j+1})}$$

where $A = (a_{ij})$ and the subscripts are taken mod n.

Here is a second instance. Let $L = (\lambda_1, \lambda_2, \ldots, \lambda_n)^T$. Then, for any permutation matrix P_σ,

$$P_\sigma(\operatorname{diag} L)P_\sigma^* = \operatorname{diag}(P_\sigma L). \tag{2.4.19}$$

A second permutation matrix of importance is

$$\Gamma = \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 & 1 \\ 0 & 0 & \ldots & 0 & 1 & 0 \\ 0 & 0 & \ldots & 1 & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 1 & \ldots & 0 & 0 & 0 \end{pmatrix}, \tag{2.4.20}$$

which corresponds to the permutation $\sigma(1) = 1$, $\sigma(2) = n$, $\sigma(3) = n - 1$, ..., $\sigma(j) = n - j + 2$, ..., $\sigma(n) = 2$. Exhibited as a product of cycles, $\sigma = (1)(2, n)(3, n - 1), \ldots, (n, 2)$. It follows that $\sigma^2 = I$, hence that

$$\Gamma^2 = I. \tag{2.4.21}$$

Also,

$$\Gamma^* = \Gamma^T = \Gamma = \Gamma^{-1}. \tag{2.4.22}$$

Again, as an instance of (2.4.13),

$$\Gamma(\operatorname{diag} L)\Gamma = \operatorname{diag}(\Gamma L). \tag{2.4.23}$$

Finally, we cite the counteridentity K, which has 1's on the main counterdiagonal and 0's elsewhere:

$$K = K_n = \begin{pmatrix} 0 & 0 & \ldots & 0 & 1 \\ 0 & 0 & \ldots & 1 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & \ldots & 0 & 0 \end{pmatrix} \tag{2.4.24}$$

One has $K = K^*$, $K^2 = I$, $K = K^{-1}$.

Let $P = P_\sigma$ designate an $n \times n$ permutation matrix.

Now σ may be factored into a product of disjoint cycles. This factorization is unique up to the arrangement of factors. Suppose that the cycles in the product have lengths $p_1, p_2, \ldots, p_m$, $(p_1 + p_2 + \cdots + p_m = n)$. Let π_{p_k} designate the π matrix (2.4.14) of order p_k. By a rearrangement of rows and columns, the cycles in P_σ can be brought into the form of involving only contiguous indices, that is, indices that are successive integers. By (2.4.13), then, there exists a permutation matrix R of order n such that

$$RPR^* = RPR^{-1} = \pi_{p_1} \oplus \pi_{p_2} \oplus \cdots \oplus \pi_{p_m}. \tag{2.4.25}$$

Since the characteristic polynomial of π_{p_k} is $(-1)^{p_k}(\lambda^{p_k} - 1)$, it follows that the characteristic polynomial of RPR*, hence of P, is $\prod_{k=1}^{m}(-1)^{p_k}(\lambda^{p_k} - 1)$. The eigenvalues of the permutation matrix P are therefore the roots of unity comprised in the totality of roots of the m equations:

$$\lambda^{p_k} = 1, \qquad k = 1, 2, \ldots, m.$$

Example. Let σ be the permutation of 1, 2, 3, 4, 5, 6 for which $\sigma(1) = 5$, $\sigma(2) = 1$, $\sigma(3) = 6$, $\sigma(4) = 4$, $\sigma(5) = 2$, $\sigma(6) = 3$. Then σ can be factored into cycles as $\sigma = (152)(4)(36)$. Therefore, $m = 3$ and $p_1 = 3$, $p_2 = 1$, $p_3 = 2$. The matrix P_σ is

$$P_\sigma = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

The matrix R corresponding to $\tau(1) = 1$, $\tau(5) = 2$, $\tau(2) = 3$, $\tau(4) = 4$, $\tau(3) = 5$, $\tau(6) = 6$, is such that

$$RP_\sigma R^* = \left(\begin{array}{ccc|c|cc} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right).$$

The eigenvalues of P_σ are therefore the roots of $(\lambda^3 - 1)(\lambda - 1)(\lambda^2 - 1)$.

A permutation σ is called primitive if its factorization consists of one cycle of full length n. The eigenvalues of a primitive permutation matrix are the nth roots of unity, hence they are distinct.

PROBLEMS

1. If M is $m \times n$, describe the relationship between M, $K_m M$, and MK_n.
2. Prove that $\det K_n = (-1)^{[n/2]}$, where [x] designates the largest integer $\leq x$.
3. Determine the characteristic polynomial of π.
4. For integer p, set $M_p = \pi^p + \pi^{-p}$. Prove that $M_p = M_{n-p}$, $M_0 = M_n = 2I$, $M_p M_q = M_{p+q} + M_{p-q}$, $M_{p+1} = M_1 M_p - M_{p-1}$.
5. Let $C_n(x) = 2 \cos n\theta$, where $x = 2 \cos \theta$, designate the Tschebyscheff polynomials of the first kind. One knows that $C_{n+1}(x) = xC_n(x) - C_{n-1}(x)$, $C_0(x) = 2$, $C_1(x) = x$. Referring to Problem 4, prove that $M_p = C_p(M_1)$.
6. Let $N = \{0, 1, 2, \ldots, 2^n - 1\}$ and let σ designate the permutation of N that results from reversing the binary bits of the elements of N.

Example. When n = 3,

$0 \to 000 \to 000 \to 0$,
$1 \to 001 \to 100 \to 4$,
$2 \to 010 \to 010 \to 2$,
$3 \to 011 \to 110 \to 6$,
$4 \to 100 \to 001 \to 1$,
$5 \to 101 \to 101 \to 5$,
$6 \to 110 \to 011 \to 3$,
$7 \to 111 \to 111 \to 7$.

Discuss the factorization of σ for n = 3. What about the general case?

7. Describe the matrices $I_m \otimes \pi_n$; $\pi_n \otimes I_m$.
8. If $m > 1$, prove that $I_m \otimes \pi_n$ and $\pi_n \otimes I_m$ are derogatory, that is, their minimal polynomial is not their characteristic polynomial.
9. Prove that $K_{2^m} \otimes K_{2^n} = K_{2^{m+n}}$.
10. Let π be of order n. Prove that $I + \pi + \pi^2 + \cdots + \pi^{n-1} = J$, where J is the matrix of all 1's.
11. If σ is a primitive permutation, prove that σ^{-1} is primitive.
12. If σ and τ are primitive permutations, is it true that $\sigma\tau$ is primitive?
13. P is a primitive permutation matrix if and only if it is of the form $P = R^* \pi R$ where R is a permutation matrix.
14. P is a primitive permutation matrix of order n if and only if n is the least positive integer for which $P^n = I$.

2.5 THE FOURIER MATRIX

Let n be a fixed integer ≥ 1 and set

$$w = \exp\left(\frac{2\pi i}{n}\right) = \cos\frac{2\pi}{n} + i \sin\frac{2\pi}{n}, \qquad i = \sqrt{-1}. \tag{2.5.1}$$

In a good deal of what follows, w might be taken as any primitive nth root of unity, but we prefer to standardize the selection as in (2.5.1). Note that

(2.5.2)
$$\begin{aligned}&\text{(a)}\quad w^n = 1,\\ &\text{(b)}\quad w\bar{w} = 1,\\ &\text{(c)}\quad \bar{w} = w^{-1},\\ &\text{(d)}\quad \bar{w}^k = w^{-k} = w^{n-k},\\ &\text{(e)}\quad 1 + w + w^2 + \cdots + w^{n-1} = 0.\end{aligned}$$

By the Fourier matrix of order n, we shall mean the matrix $F\ (= F_n)$ where

(2.5.3)
$$F^* = \frac{1}{\sqrt{n}}(w^{(i-1)(j-1)})$$

$$= \frac{1}{\sqrt{n}}\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & w & w^2 & \cdots & w^{n-1} \\ 1 & w^2 & w^4 & \cdots & w^{2(n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & w^{n-1} & w^{2(n-1)} & \cdots & w^{(n-1)(n-1)} \end{pmatrix}.$$

Note the star on the left-hand member. The sequence w^k, $k = 0, 1, \ldots$, is periodic; hence there are only n distinct elements in F. F can therefore be written alternatively as

(2.5.4)
$$F^* = n^{-1/2}\begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & w & w^2 & \cdots & w^{n-1} \\ 1 & w^2 & w^4 & \cdots & w^{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & w^{n-1} & w^{n-2} & \cdots & w \end{pmatrix}.$$

It is easily established that F and F* are symmetric:

(2.5.5)
$$F = F^T,\quad F^* = (F^*)^T = \bar{F},\quad F = \bar{F}^*.$$

It is of fundamental importance that

Theorem 2.5.1. F is unitary:

(2.5.6)
$$FF^* = F^*F = I \quad \text{or} \quad F^{-1} = F^* \quad \text{or}$$
$$F\bar{F}^T = \bar{F}^TF = I \quad \text{or} \quad F^{-1} = \bar{F}^T.$$

Proof. This is a result of the geometric series identity

$$\sum_{r=0}^{n-1} w^{r(j-k)} = \frac{1 - w^{n(j-k)}}{1 - w^{j-k}} = \begin{cases} n \text{ if } j = k, \\ 0 \text{ if } j \neq k. \end{cases}$$

A second application of the geometrical identity yields

Theorem 2.5.2

$$F^{*2} = F^*F^* = \Gamma = \begin{pmatrix} 1 & 0 & \cdots & & 0 \\ 0 & 0 & \cdots & & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 1 & \cdots & & 0 \end{pmatrix} = F^{\lambda}$$

Corollary. $F^{*4} = \Gamma^2 = I$. $F^{*3} = F^{*4}(F^*)^{-1} = IF = F$.

We may write the Fourier matrix picturesquely in the form

(2.5.7)
$$F = \sqrt[4]{I}.$$

(It may be shown that all the qth roots of I are of the form $M^{-1}DM$ where $D = \operatorname{diag}(\mu_1, \mu_2, \ldots, \mu_n)$, $\mu_i^q = 1$ and where M is any nonsingular matrix.)

Corollary. The eigenvalues of F are ± 1, $\pm i$, with appropriate multiplicities.

Carlitz has obtained the characteristic polynomials $f(\lambda)$ of F^* $(= F_n^*)$. They are as follows.

$$n \equiv 0 \pmod 4, \qquad f(\lambda) = (\lambda - 1)^2(\lambda + i)(\lambda + 1)(\lambda^4 - 1)^{(n/4)-1},$$

$$\begin{aligned} &n \equiv 1 \pmod 4, && f(\lambda) = (\lambda - 1)(\lambda^4 - 1)^{(1/4)(n-1)}, \\ &n \equiv 2 \pmod 4, && f(\lambda) = (\lambda^2 - 1)(\lambda^4 - 1)^{(1/4)(n-2)}, \\ &n \equiv 3 \pmod 4, && f(\lambda) = (\lambda + i)(\lambda^2 - 1)(\lambda^4 - 1)^{(1/4)(n-3)}. \end{aligned}$$

The discrete Fourier transform. Working with complex n-tuples, write

$$Z = (z_1, z_2, \ldots, z_n)^T \quad \text{and}$$
$$\hat{Z} = (\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_n)^T.$$

The linear transformation

$$\hat{Z} = FZ \tag{2.5.8}$$

where F is the Fourier matrix is known as the discrete Fourier transform (DFT). Its inverse is given simply by

$$Z = F^{-1}\hat{Z} = F^*\hat{Z}. \tag{2.5.9}$$

The transform (2.5.8) often goes by the name of harmonic analysis or periodogram analysis, while the inverse transform (2.5.9) is called harmonic synthesis. The reasons behind these terms are as follows: suppose that $p(z) = a_0 + a_1 z + \cdots + a_n z^{n-1}$ is a polynomial of degree $\leq n - 1$. It will be determined uniquely by specifying its values $p(z_k)$ at n distinct points z_k, $k = 1, 2, \ldots, n$ in the complex plane. Select these points z_k as the n roots of unity $1, w, w^2, \ldots, w^{n-1}$. Then clearly

$$n^{1/2}F^* \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{pmatrix} = \begin{pmatrix} p(1) \\ p(w) \\ \vdots \\ p(w^{n-1}) \end{pmatrix} \tag{2.5.10}$$

so that

$$\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{n-1} \end{pmatrix} = n^{-1/2}F \begin{pmatrix} p(1) \\ p(w) \\ \vdots \\ p(w^{n-1}) \end{pmatrix}. \tag{2.5.11}$$

The passage from functional values to coefficients through (2.5.11) or (2.5.8) is an analysis of the function, while in the passage from coefficient values to functional values through (2.5.10) or (2.5.9) the functional values are built up or "synthesized."

These formulas for interpolation at the roots of unity can be given another form.

By a Vandermonde matrix $V(z_0, z_1, \ldots, z_{n-1})$ is meant a matrix of the form

$$V = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ z_0 & z_1 & \cdots & z_{n-1} \\ z_0^2 & z_1^2 & \cdots & z_{n-1}^2 \\ \vdots & \vdots & & \vdots \\ z_0^{n-1} & z_1^{n-1} & \cdots & z_{n-1}^{n-1} \end{pmatrix} \tag{2.5.12}$$

From (2.5.4) one has, clearly,

$$\begin{aligned} &V(1, w, w^2, \ldots, w^{n-1}) = n^{1/2}F^*, \\ &V(1, \bar{w}, \bar{w}^2, \ldots, \bar{w}^{n-1}) = n^{1/2}\bar{F}^* = n^{1/2}F. \end{aligned} \tag{2.5.13}$$

One now has from (2.5.11)

$$\begin{aligned} p(z) &= (1, z, \ldots, z^{n-1})(a_0, a_1, \ldots, a_{n-1})^T \\ &= (1, z, \ldots, z^{n-1})n^{-1/2} \\ &\quad F(p(1), p(w), \ldots, p(w^{n-1}))^T \\ &= n^{-1/2}(1, z, \ldots, z^{n-1})V(1, \bar{w}, \bar{w}^2, \\ &\quad \ldots, \bar{w}^{n-1})(p(1), p(w), \ldots, p(w^{n-1}))^T. \end{aligned} \tag{2.5.14}$$

Note. In the literature of signal processing, a sequence-to-sequence transform is known as a discrete or digital filter. Very often the transform [such as (2.5.8)] is linear and is called a linear filter.

Fourier Matrices as Kronecker Products. The Fourier matrices of orders 2^n may be expressed as Kronecker products. This factorization is a manifestation, essentially, of the idea known as the Fast Fourier Transform (FFT) and is of vital importance in real time calculations.

Let F'_{2^n} designate the Fourier matrices of order 2^n whose rows have been permuted according to the bit reversing permutation (see Problem 6, p. 30).

Examples

$$F'_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

$$F'_4 = \frac{1}{\sqrt{4}} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \\ 1 & -i & -1 & i \end{pmatrix}.$$

One has

$$F'_4 = (I_2 \otimes F'_2)D_4(F'_2 \otimes I_2), \tag{2.5.15}$$

where $D_4 = \mathrm{diag}(1, 1, 1, i)$. This may be easily checked out.

As is known, $A \otimes B = P(B \otimes A)P^*$ for some permutation matrix P that depends merely on the dimensions of A and B. We may therefore write, for some permutation matrix S_4 (one has, in fact, $S_4^{-1} = S_4$):

$$\boxed{F'_4 = (I_2 \otimes F'_2)D_4S_4(I_2 \otimes F'_2)S_4.} \tag{2.5.16}$$

Similarly,

$$F'_{16} = (I_4 \otimes F'_4)D_{16}(F'_4 \otimes I_4) \tag{2.5.17}$$

where

$$D_{16} = \mathrm{diag}(I, D^2, D, D^3) \tag{2.5.18}$$

with

$$D = \mathrm{diag}(1, w, w^2, w^3), \qquad w = \exp\frac{2\pi i}{16}. \tag{2.5.19}$$

Again, for an appropriate permutation matrix $S_{16} = S_{16}^{-1} = S_{16}^T$,

$$\boxed{F'_{16} = (I_4 \otimes F'_4)D_{16}S_{16}(I_4 \otimes F'_4)S_{16}.} \tag{2.5.20}$$

For 256 use

$$D_{256} = \mathrm{diag}(I, D^8, D^4, \ldots, D^{15}) \tag{2.5.21}$$

where the sequence 0, 8, 4, ..., 15 is the bit reversed order of 0, 1, ..., 15 and where

$$D = \mathrm{diag}(1, w, \ldots, w^{15}), \qquad w = e^{2\pi i/256}. \tag{2.5.22}$$

PROBLEMS

1. Evaluate $\det F_n$.
2. Find the polynomial $p_{n-1}(z)$ of degree $\leq n - 1$ that takes on the values $1/z$ at the nth roots of unity, w^j, $j = 1, 2, \ldots, n$. What is the limiting behavior of $p_n(z)$ as $n \to \infty$? (de Méré)
3. Write $F = R + iS$ where R and S are real and $i = \sqrt{-1}$. Show that R and S are symmetric and that $R^2 + S^2 = I$, $RS = SR$.
4. Exhibit R and S explicitly.

2.6 HADAMARD MATRICES

By a Hadamard matrix of order n, H ($= H_n$), is meant a matrix whose elements are either $+1$ or -1 and for which

(2.6.1) $$\boxed{HH^T = H^TH = nI.}$$

Thus, $n^{-1/2}H$ is an orthogonal matrix.

Examples

$$H_1 = (1),$$

$$\sqrt{2}\,F_2 = H_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

$$H_{4,1} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 \end{pmatrix},$$

$$H_{4,2} = \begin{pmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ -1 & 1 & 1 & 1 \end{pmatrix}.$$

It is known that if $n > 3$, then the order of an Hadamard matrix must be a multiple of 4. With one possible exception, all multiples of $4 \leq 200$ yield at least one Hadamard matrix.

Theorem 2.6.1. If A and B are Hadamard matrices of orders m and n respectively, then $A \otimes B$ is an Hadamard matrix of order mn.

Proof

$$(A \otimes B)(A \otimes B)^T = (A \otimes B)(A^T \otimes B^T) = (AA^T) \otimes (BB^T)$$

$$= (mI_m) \otimes (nI_n) = mn(I_m \otimes I_n) = mnI_{mn}.$$

In some areas, particularly digital signal processing, the term Hadamard matrix is limited to the matrices of order 2^n given specifically by the recursion

(2.6.2) $$\boxed{\begin{aligned} H_1 &= (1), \quad H_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \\ H_{2^{n+1}} &= H_{2^n} \otimes H_{2^n}. \end{aligned}}$$

These matrices have the additional property of being symmetric,

(2.6.3) $$H_{2^n} = H^T_{2^n},$$

so that

(2.6.4) $$H^2_{2^n} = 2^nI.$$

The Walsh-Hadamard Transform. By this is meant the transform

(2.6.5) $$\boxed{\hat{Z} = HZ}$$

where H is an Hadamard matrix.

PROBLEMS

1. Hadamard parlor game: Write down in a row any four numbers. Then write the sum of the first two, the sum of the last two; the difference of the first two, the difference of the last two to form a second row. Iterate this procedure four times. The final row will be four times the original row. Explain, making reference to H_4. Generalize.
2. Define a generalized permutation matrix P as follows. P is square and every row and every column of P has exactly one nonzero element in it. That element is either a +1 or a -1. Show that if H is an Hadamard matrix, and if P and Q are generalized permutation matrices, then PHQ is an Hadamard matrix.
3. With the notation of (2.6.2) prove that

$$H_{2^{n+1}} = (H_{2^n} \otimes I_2)(I_{2^n} \otimes H_2).$$

4. Using Problem 3, show that the Hadamard transform of a vector by H_{2^n} can be carried out in $\leq n\,2^n$ additions or subtractions.
5. If H is an Hadamard matrix of order n, prove that $|\det H| = n^{n/2}$.

2.7 TRACE

The trace of a square matrix $A = (a_{ij})$ of order n is defined as the sum of its diagonal elements:

$$\text{(2.7.1)} \qquad \boxed{\operatorname{tr} A = \sum_{j=1}^{n} a_{jj}.}$$

The principal general properties of the trace are

(1) $\operatorname{tr}(aA + bB) = a \operatorname{tr}(A) + b \operatorname{tr}(B)$.

(2) $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

(3) $\operatorname{tr} A = \operatorname{tr}(S^{-1}AS)$, S nonsingular.

(4) If λ_i are the eigenvalues of A, then $\operatorname{tr} A = \sum_{i=1}^{n} \lambda_i$.

(5) More generally, if p designates a polynomial

$$p(\lambda) = \sum_{j=0}^{r} a_j \lambda^j,$$

then $\operatorname{tr}(p(A)) = \sum_{k=1}^{n} p(\lambda_k)$.

(6) $\operatorname{tr}(AA^*) = \operatorname{tr}(A^*A) = \sum_{i,j=1}^{n} |a_{ij}|^2$ = square of Frobenius norm of A.

(7) $\operatorname{tr}(A \oplus B) = \operatorname{tr} A + \operatorname{tr} B$.

(8) $\operatorname{tr}(A \otimes B) = (\operatorname{tr} A)(\operatorname{tr} B)$.

2.8 GENERALIZED INVERSE

For large classes of matrices, such as the square "singular" matrices and the rectangular matrices, no

inverse exists. That is, there are many matrices A for which there exists no matrix B such that AB = BA = I.

In discussing the solution of systems of linear equations, we know that if A is n × n and nonsingular then the solution of the equation

$$AX = B,$$

where X and B are n × m matrices, can be written very neatly in matrix form as

$$\boxed{X = A^{-1}B.}$$

Although the "solution" give above is symbolic, and in general is not the most economical way of solving systems of linear equations, it has important applications. However, we have so far only been able to use this idea for square nonsingular matrices. In this section we show that for every matrix A, whether square or rectangular, singular or nonsingular, there exists a unique "generalized inverse" often called the "Moore-Penrose" inverse of A, and employing it, the formal solution $X = A^{-1}B$ can be given a useful interpretation. This generalized inverse has several of the important properties of the inverse of a square nonsingular matrix, and the resulting theory is able in a remarkable way to unify a variety of diverse topics. This theory originated in the 1920s, but was rediscovered in the 1950s and has been developed extensively since then.

2.8.1 Right and Left Inverses

Definition. If A is an m × n matrix, a right inverse of A is an n × m matrix B such that $AB = I_m$. Similarly a left inverse is a matrix C such that $CA = I_n$.

Example. If

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix},$$

a right inverse of A is the matrix

$$B = \begin{pmatrix} 2 & -1 \\ -1 & 1 \\ 0 & 0 \end{pmatrix},$$

since $AB = I_2$.

However, note that A does not have a left inverse, since for any matrix C, by the theorem on the rank of a product, $r(CA) \leq r(A) = 2$, so that $CA \neq I_3$. Similarly, although A is, by definition, a left inverse of B, there exists no right inverse of B.

The following theorem gives necessary and sufficient conditions for the existence of a right or left inverse.

Theorem 2.8.1.1. An $m \times n$ matrix A has a right (left) inverse if and only if A has rank m(n).

Proof. We work first with right inverses. Assume that $AB = I_m$. Then $m = r(I_m) \leq r(A) \leq m$. Hence $r(A) = m$.

Conversely, suppose that $r(A) = m$. Then A has m linearly independent columns, and we can find a permutation matrix P so that the matrix $\hat{A} = AP$ has its first m columns linearly independent. Now, if we can find a matrix $\hat{B}$ such that $\hat{A}\hat{B} = AP\hat{B} = I$, then $B = P\hat{B}$ is clearly a right inverse for A.

Therefore, we may assume, without loss of generality, that A has its first m columns linearly independent. Hence A can be written in the block form

$$A = (A_1, A_2)$$

where A_1 is an $m \times m$ nonsingular matrix and A_2 is some $m \times (n - m)$ matrix. This can be factored to yield

$$A = A_1(I_m, Q) \qquad (Q = A_1^{-1}A_2).$$

Now let

$$B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}$$

where B_1 is $m \times n$ and B_2 is $(n - m) \times m$. Then $AB = I$

if and only if

$$A_1B_1 + A_1QB_2 = I,$$

or if and only if

$$B_1 + QB_2 = A_1^{-1},$$

or if and only if

$$B_1 = A_1^{-1} - QB_2.$$

Therefore, we have

$$B = \begin{pmatrix} A_1^{-1} - QB_2 \\ \hline B_2 \end{pmatrix} = \begin{pmatrix} A_1^{-1} \\ 0 \end{pmatrix} - \begin{pmatrix} Q \\ -I \end{pmatrix} B_2 .$$

for an arbitrary $(n - m) \times m$ matrix B_2. Thus there is a right inverse, and if $n > m$, it is not unique.

We now prove the theorem for a left inverse. Suppose, again, that A is $m \times n$ and $r(A) = n$. Then A^T is $n \times m$ and $r(A^T) = n$. By the first part, A^T has a right inverse: $A^TB = I$. Hence $B^TA = I$ and A has a left inverse.

Corollary. If A is $n \times n$ of rank n, then A has both a right and a left inverse and they are the same.

Proof. The existence of a right and a left inverse for A follows immediately from the theorem. To prove that they are the same we assume

$$AB = I, \qquad CA = I.$$

Then $C(AB) = CI = C$. But also,

$$C(AB) = (CA)B = IB = B,$$

so that $B = C$. This is the matrix that is defined to be the inverse of A, denoted by A^{-1}.

PROBLEMS

1. Find a left inverse for $\begin{pmatrix} 1 & 1 \\ 2 & 0 \\ 3 & 1 \end{pmatrix}$. Find all the left inverses.

2. Does

$$\begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 4 \\ 1 & 3 & 4 \\ 1 & 4 & 5 \end{pmatrix}$$

have a left inverse?

3. Let A be m × n and have a left inverse B. Suppose that the system of linear equations AX = C has a solution. Prove that the solution is unique and is given by X = BC.

4. Let B be a left inverse for A. Prove that ABA = A and BAB = B.

5. Let A be m × n and have rank n. Prove that A^TA is nonsingular and that $(A^TA)^{-1}A^T$ is a left inverse for A.

6. Let A be m × n and have rank n. Let W be m × m positive definite symmetric. Prove that A^TWA is nonsingular and that $(A^TWA)^{-1}A^TW$ is a left inverse for A.

2.8.2 Generalized Inverses

Definition. Let A be an m × n matrix. Then an n × m matrix X that satisfies any or all of the following properties is called a generalized inverse:

(1) AXA = A,

(2) XAX = X,

(3) (AX)* = AX,

(4) (XA)* = XA.

Here the star * represents the conjugate transpose. A matrix satisfying all four of the properties above is called a Moore-Penrose inverse of A (for short: an M-P inverse). We show now that every matrix A has a unique M-P inverse. It is denoted by $A^{\dagger}$. It should be remarked that the M-P inverse is often designated

by other symbols, such as A^{+}. The notation $A^{\dagger}$ is used here because (a) it is highly suggestive and (b) it comes close to one used in the APL computer language.

We first prove the following lemma on "rank factorization" of a matrix.

Lemma. If A is an m × n matrix of rank r, then A = BC, where B is m × r, C is r × n and r(B) = r(C) = r.

Proof. Since the rank of A is r, A has r linearly independent columns. We may assume, without loss of generality, that these are the first r columns of A, for, if not, there exists a permutation matrix P such that the first r columns of the matrix AP are the r linearly independent columns of A. But if AP can be factored as

$$AP = B\hat{C}, \qquad r(B) = r(\hat{C}) = r,$$

then

$$A = BC$$

where $C = \hat{C}P^{-1}$ and $r(C) = r(\hat{C}) = r$, since P is nonsingular.

Thus if we let B be the m × r matrix consisting of the first r columns of A, the remaining n - r columns are linear combinations of the columns of B, of the form $BQ^{(j)}$ for some r × 1 vector $Q^{(j)}$. Then if we let Q be the r × (n - r) matrix,

$$Q = (Q^{(1)} \cdots Q^{(n-r)}),$$

we have

$$A = \overset{r}{(B,} \ \overset{n-r}{BQ)}$$

(letters over blocks indicate number of columns).

If we let

$$C = (I_r, Q),$$

we have

$$A = B(I_r, Q) = BC$$

and r(B) = r(C) = r.

We next show the existence of an M-P inverse in the case where A has full row or full column rank.

Theorem 2.8.2.1

(a) If A is square and nonsingular, set $A^{\dagger} = A^{-1}$.

(b) If A is $n \times 1$ (or $1 \times n$) and $A \neq 0$, set

$$A^{\dagger} = \frac{1}{(A^*A)} A^* \qquad (\text{or } A^{\dagger} = \frac{1}{(AA^*)} A^*).$$

(c) If A is $m \times n$ and $r(A) = m$, set $A^{\dagger} = A^*(AA^*)^{-1}$. If A is $m \times n$ and $r(A) = n$, set $A^{\dagger} = (A^*A)^{-1}A^*$.

Then $A^{\dagger}$ is an M-P inverse for A. Moreover, in the case of full row rank, it is a right inverse; in the case of full column rank, it is a left inverse.

Note that (a) and (b) are really special cases of (c).

Proof. Direct calculation. Observe that if A is $m \times n$ and $r(A) = m$, then AA^* is $m \times m$. It is well known that $r(AA^*) = m$, so that $(AA^*)^{-1}$ can be formed. Similarly for A^*A.

We can now show the existence of an M-P inverse for any $m \times n$ matrix A.

If $A = 0$, set $A^{\dagger} = 0^* = 0_{n,m}$. This is readily verified to satisfy requirements (1), (2), (3) and (4) for a generalized inverse.

If $A \neq 0$, factor A as in the lemma into the product

$$A = BC$$

where B is $m \times r$, C is $r \times n$ and $r(B) = r(C) = r$. Now B has full column rank while C has full row rank, so that $B^{\dagger}$ and $C^{\dagger}$ may be found as in the previous theorem. Now set

$$A^{\dagger} = C^{\dagger}B^{\dagger}.$$

Theorem 2.8.2.2. Let $A^{\dagger}$ be defined as above. Then it is an M-P inverse for A.

$A^{+} = (E A^{T}A)^{+}(EA^{T})$

Proof. It is easier to verify properties (3) and (4) first. They will then be used in proving properties (1) and (2).

(3) $AA^{\dagger} = B(CC^{\dagger})B^{\dagger} = BIB^{\dagger} = BB^{\dagger}$, and since $BB^{\dagger} = (BB^{\dagger})^*$, we have $AA^{\dagger} = (AA^{\dagger})^*$.

(4) Similarly, $A^{\dagger}A = C^{\dagger}C = (C^{\dagger}C)^* = (A^{\dagger}A)^*$.

(1) $(AA^{\dagger})A = (BB^{\dagger})BC = BC = A$.

(2) $(A^{\dagger}A)A^{\dagger} = (C^{\dagger}C)C^{\dagger}B^{\dagger} = C^{\dagger}B^{\dagger} = A^{\dagger}$.

Now we prove that for any matrix A the M-P inverse is unique.

Theorem 2.8.2.3. Given an $m \times n$ matrix A, there is only one matrix $A^{\dagger}$ that satisfies all four properties for the Moore-Penrose inverse.

Proof. Suppose that there exist matrices B and C satisfying

$ABA = A$	(1)	$ACA = A$,
$BAB = B$	(2)	$CAC = C$,
$(AB)^* = AB$	(3)	$(AC)^* = AC$,
$(BA)^* = BA$	(4)	$(CA)^* = CA$.

Then

$$B \overset{(2)}{=} (BA)B \overset{(4)}{=} (A^*B^*)B \overset{(1)}{=} (A^*C^*A^*)B^*B$$

$$\overset{(4)}{=} (CA)(A^*B^*B) \overset{(4)}{=} CA(BAB) \overset{(2)}{=} CAB$$

and

$$C \overset{(2)}{=} C(AC) \overset{(3)}{=} CC^*A^* \overset{(1)}{=} CC^*(A^*B^*A^*)$$

$$\overset{(3)}{=} (CC^*A^*)(AB) \overset{(3)\text{ and }(2)}{=} CAB.$$

Therefore $B = C$. The integers over the equality signs show the equations used to derive the equality.

Penrose has given the following recursive method

for computing $A^{\dagger}$, which is included in case the reader would like to write a computer program.

Theorem 2.8.2.4 (the Penrose algorithm). Let A be $m \times n$ and have rank $r > 0$.

(a) Set $B = A^*A$ (B is $n \times n$).

(b) Set $C_1 = I$ (C_1 is $n \times n$).

(c) Set recursively for $i = 1, 2, \ldots, r-1$:
$C_{i+1} = (1/i)\,\mathrm{tr}(C_iB)I - C_iB$ (C_i is $n \times n$).

Then $\mathrm{tr}(C_rB) \neq 0$ and $A^{\dagger} = rC_rA^*/\mathrm{tr}(C_rB)$. Moreover, $C_{r+1}B = 0$. We therefore do not need to know r beforehand, but merely stop the recurrence when we have arrived at this stage.

The proof is omitted.

Also very useful is the Greville algorithm.

Theorem 2.8.2.5. Define $A_k = (A_{k-1}\ a_k)$ where a_k is the kth column of A and A_{k-1} is the submatrix of A consisting of its first $k-1$ columns. Set $d_k = A_{k-1}^{\dagger}a_k$ and $c_k = a_k - A_{k-1}d_k$. Set $b_k = c_k^{\dagger}$ if $c_k \neq 0$. If $c_k = 0$, set $b_k = (1 + d_k^*d_k)^{-1}d_k^*A_{k-1}^{\dagger}$. Then

$$A^{\dagger} = A_k^{\dagger} = \begin{pmatrix} A_{k-1}^{\dagger} - d_kb_k \\ b_k \end{pmatrix}.$$

To start: set $A_1^{\dagger} = 0$ if $a_1 = 0$; if not, set $A_1^{\dagger} = (a_1^*a_1)^{-1}a_1^*$.

PROBLEMS

1. If $A = \begin{pmatrix} 2 & 2 & 0 \\ 1 & 2 & 1 \\ 1 & 2 & 1 \end{pmatrix}$, verify that

$$A^{\dagger} = \frac{1}{24}\begin{pmatrix} 18 & 3 & -5 \\ -6 & 3 & 9 \\ -6 & 3 & 9 \end{pmatrix}.$$

2. If $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}$, find $A^{\dagger}$.

3. If

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{pmatrix},$$

find $A^{\dagger}$.

4. Use Penrose's formulas to compute the inverse of the nonsingular matrix

$$\begin{pmatrix} 14 & 8 & 3 \\ 8 & 5 & 2 \\ 3 & 2 & 1 \end{pmatrix}.$$

Use Greville's algorithm.

5. If c is a nonzero scalar, prove that $(cA)^{\dagger} = (1/c)A^{\dagger}$.

6. Prove that $(A^{\dagger})^{\dagger} = A$.

7. Prove that $(A^{\dagger})^* = (A^*)^{\dagger}$.

8. If d is a scalar, define $d^{\dagger}$ by $d^{\dagger} = d^{-1}$ if $d \neq 0$, $d^{\dagger} = 0$ if $d = 0$. Let $A = \mathrm{diag}(d_1, \ldots, d_n)$. Prove that $A^{\dagger} = \mathrm{diag}(d_1^{\dagger}, \ldots, d_n^{\dagger})$.

9. Prove that $\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}^{\dagger} = \begin{pmatrix} A^{\dagger} & 0 \\ 0 & B^{\dagger} \end{pmatrix}$ and $\begin{pmatrix} 0 & A \\ B & 0 \end{pmatrix}^{\dagger} = \begin{pmatrix} 0 & B^{\dagger} \\ A^{\dagger} & 0 \end{pmatrix}$.

10. Prove that if $A^{\dagger} = 0$, then $A = 0^*$.

11. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and have rank 1. Prove that

$$A^{\dagger} = \frac{1}{|a|^2 + |b|^2 + |c|^2 + |d|^2}A^*.$$

12. Let J be the J matrix of order n. Prove that $J^{\dagger} = (1/n^2)J$.

13. Let S be an $n \times n$ matrix with 1's on the superdiagonal and 0's elsewhere. Find $S^{\dagger}$.

14. Let P be any projection matrix (i.e., $P^2 = P$, $P^* = P$). Prove that $P^{\dagger} = P$.

15. Prove that both $AA^{\div}$ and $A^{\div}A$ are projections.
16. Prove that $A^{\div} = (A^*A)^{\div}A^* = A^*(AA^*)^{\div}$.
17. Prove that $r(A) = r(A^{\div}) = r(A^{\div}A) = r(A^*A^{\div})$.
18. Taking $A = (1, 0)$, $B = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, show that, in general, $(AB)^{\div} \neq B^{\div}A^{\div}$.
19. If a and b are column vectors, then $a^{\div} = (a^*a)^{\div}a^*$, and $(ab^*)^{\div} = (a^*a)^{\div}(b^*b)^{\div}ba^*$.
20. Prove that $(A \otimes B)^{\div} = A^{\div} \otimes B^{\div}$.

2.8.3 The UDV Theorem and the M-P Inverse

We begin by establishing a theorem that is of great utility in visualizing the action and facilitating the manipulation of rectangular (or square) matrices. This is the UDV theorem, also called the diagonal decomposition theorem or the singular value decomposition theorem.

Theorem 2.8.3.1. Let A be an $m \times n$ matrix with complex elements and of rank r. Then the exist unitary matrices U, V of orders m and n respectively such that

$$A = UDV^* \tag{2.8.3.1}$$

where

$$D = \begin{pmatrix} D_1 & 0 \\ 0 & 0 \end{pmatrix} \tag{2.8.3.2}$$

is $m \times n$ and where $D_1 = \text{diag}(d_1, d_2, \ldots, d_r)$ is a nonsingular diagonal matrix of order r.

Note that the representation (2.8.3.1) can be written as $U^*AV = D$ or, changing the notation, $UAV = D$, and so on (since U and V are unitary).

Let A be $m \times n$; then, as is well known, AA^* is positive semidefinite Hermitian ~~symmetric~~ and $r(AA^*) = r(A) = r(A^*)$. Hence the eigenvalues of AA^* are real and nonnegative. Write them as $d_1^2, d_2^2, \ldots, d_r^2, 0, 0, \ldots, 0$ where the d_i's are positive and where there are $m - r$ 0's in the list. The numbers $d_1, d_2, \ldots, d_r$ are known as the singular values of A.

Proof. Define $D_1 = \text{diag}(d_1, d_2, \ldots, d_r)$. Let U_1 be $m \times r$ and consist of the (orthonormal) eigenvectors of AA^* corresponding to the eigenvalues $d_1^2, d_2^2, \ldots, d_r^2$ (cf. Theorems 2.9.3 and 2.9.9). We have $AA^*U_1 = U_1D_1^2$ and $U_1^*U_1 = I_r$. Let U_2 be the $m \times (m - r)$ matrix whose columns consist of an orthonormal basis for the null space of A^*. Then $A^*U_2 = 0$ and $U_2^*U_2 = I_{m-r}$.

Write $U = (U_1, U_2)$ (block notation). Then

$$U^*U = \begin{pmatrix} U_1^* \\ U_2^* \end{pmatrix} (U_1, U_2) = \begin{pmatrix} U_1^*U_1 & U_1^*U_2 \\ U_2^*U_1 & U_2^*U_2 \end{pmatrix}.$$

Now, since $AA^*U_1 = U_1D_1^2$, $U_2^*AA^*U_1 = U_2^*U_1D_1^2$. But $A^*U_2 = 0$, so that $U_2^*A = 0$, hence $U_2^*U_1D_1^2 = 0$. Since D_1^2 is nonsingular, it follows that $U_2^*U_1 = U_1^*U_2 = 0$. This means that

$$U^*U = \begin{pmatrix} I_r & 0 \\ 0 & I_{m-r} \end{pmatrix} = I_m,$$

and hence that U is unitary.

Let V_1 be the $n \times r$ matrix defined by $V_1 = A^*U_1D_1^{-1}$. Let V_2 be the $n \times (n - r)$ matrix whose $n - r$ columns are a set of $n - r$ orthonormal vectors for the null space of A. Thus $AV_2 = 0$ and $V_2^*V_2 = I_{n-r}$. Define V as the $n \times n$ matrix $V = (V_1, V_2)$. Now

$$V_1^*V_1 = (D_1^{-1}U_1^*A)(A^*U_1D_1^{-1}) = D_1^{-1}U_1^*U_1D_1^2D_1^{-1}$$
$$= D_1^{-1}D_1 = I_r,$$

and $V_2^*V_1 = V_2^*A^*U_1D_1^{-1} = (AV_2)^*U_1D_1^{-1} = 0$. It follows that V is unitary. Finally,

$$U^*AV = \begin{pmatrix} U_1^* \\ U_2^* \end{pmatrix} A(V_1, V_2) = \begin{pmatrix} U_1^*AV_1 & U_1^*AV_2 \\ U_2^*AV_1 & U_2^*AV_2 \end{pmatrix}$$

$$= \begin{pmatrix} U_1^*AV_1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} U_1^*AA^*U_1D_1^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} D_1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Using UDV theorem, we can produce a very convenient formula for $A^\dagger$.

Theorem 2.8.3.2. If $A = U^*DV^*$, where U, V, D are as above, then

$$\boxed{A^\dagger = VD^\dagger U}$$

where

$$D^\dagger = \begin{array}{c} \begin{matrix} r & m-r \end{matrix} \\ \begin{pmatrix} D_1^{-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{matrix} r \\ n-r \end{matrix} \end{array}.$$

Proof. By a direct computation, it is easy to show that the n × m matrix

$$\begin{array}{c} \begin{matrix} r & m-r \end{matrix} \\ \begin{pmatrix} D_1^{-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{matrix} r \\ n-r \end{matrix} \end{array}$$

is $D^\dagger$. Now since $A(VD^\dagger U) = U^*DD^\dagger U = U^*\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}U$ and $(VD^\dagger U)A = V\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}V^*$, the third and fourth properties for the generalized inverse are satisfied. Also, $AA^\dagger A = (U^*DV^*)(VD^\dagger U)(U^*DV^*) = U^*DD^\dagger DV^* = U^*DV^* = A$. Similarly $A^\dagger AA^\dagger = A^\dagger$, proving the first two properties.

Theorem 2.8.3.3. For each A there exist polynomials p and q such that

$$A^\dagger = A^*p(AA^*),$$

$$A^\dagger = q(A^*A)A^*.$$

Proof. Let A be m × n and have rank r. Then by the diagonal decomposition theorem there exist unitary matrices U, V of order m and n and an m × n matrix

$$D = \begin{array}{c} \begin{matrix} r & n-r \end{matrix} \\ \begin{pmatrix} D_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{matrix} r \\ m-r \end{matrix} \end{array},$$

where $D_1 = \operatorname{diag}(d_1, d_2, \ldots, d_r)$, $d_1d_2\cdots d_r \neq 0$, such that $A = U^*DV^*$. Then $A^* = VD^*U$, $AA^* = U^*DD^*U$, and $A^\dagger = VD^\dagger U$. For an arbitrary polynomial $p(z)$, $p(AA^*) = p(U^*(DD^*)U) = U^*p(DD^*)U$. Hence $A^*p(AA^*) = VD^*p(DD^*)U$. Therefore for $A^\dagger$ to equal $A^*p(AA^*)$ it is necessary and sufficient that $D^\dagger = D^*p(DD^*)$. Equivalently,

$$\begin{pmatrix} D_1^\dagger & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} D_1^* & 0 \\ 0 & 0 \end{pmatrix} p\begin{pmatrix} D_1D_1^* & 0 \\ 0 & 0 \end{pmatrix}$$

or $d_k^{-1} = \bar{d}_k p(|d_k|^2)$, $k = 1, 2, \ldots, r$. Thus $p(|d_k|^2) = 1/(|d_k|^2)$, $k = 1, 2, \ldots, r$ is necessary and sufficient. Let s designate the number of distinct values among $|d_1|, |d_2|, \ldots, |d_r|$. Then by the fundamental theorem of polynomial interpolation (see any book on interpolation, approximation, or numerical analysis) there is a unique polynomial of degree $\leq s - 1$ that takes on the values $|d_k|^{-2}$ at the s points $|d_k|^2$.

The second identity for $A^\dagger$ is proved similarly.

PROBLEMS

1. Let U and V be unitary. Prove that $(UAV)^\dagger = V^*A^\dagger U^*$.
2. Let A be normal. Give a representation for $A^\dagger$ in terms of the characteristic values of A. See Section 2.9.
3. Prove that if A is normal, $AA^\dagger = A^\dagger A$.
4. Prove that $A^\dagger = A^*$ if and only if the singular

values of A are 0 or 1.

5. Prove that $A^{\dagger} = \lim_{t \to 0} A^*(tI + AA^*)^{-1}$.

2.8.4 Generalized Inverses and Systems of Linear Equations

Using the properties of the generalized inverse we are able to determine, for any system of equations

$$AX = B,$$

whether or not the system has a solution. If it does, we can obtain a matrix equation, involving the generalized inverse, which exhibits this solution. Oddly enough, we need only the first property of a generalized inverse. That is, we may use any matrix $A^{(1)}$, such that $AA^{(1)}A = A$.

Definition. If A is $m \times n$, any $n \times m$ matrix $A^{(1)}$ that satisfies $AA^{(1)}A = A$ is called a (1)-inverse of A. More generally, any matrix that satisfies any combination of the four requirements for the generalized inverse on page 44 is designated accordingly.

Example. A (1, 2, 4)-inverse for A is one that satisfies conditions (1), (2), and (4).

Theorem 2.8.4.1. Let A be $m \times n$. The system of equations

$$AX = B$$

has a solution if and only if $B = AA^{(1)}B$, for any (1)-inverse $A^{(1)}$ of A. In this case, the general solution is given by

$$X = A^{(1)}B + (I - A^{(1)}A)Y$$

for an arbitrary $n \times 1$ vector Y.

Proof. Let $B = AA^{(1)}B$. Then $AX = AA^{(1)}B$ is solved by $X = A^{(1)}B$. Suppose, conversely, that the system has a solution X_0: $AX_0 = B$. Then, for any

(1)-inverse, $A^{(1)}$,

$$B = AX_0 = (AA^{(1)}A)X_0 = AA^{(1)}B.$$

Moreover, if $X = A^{(1)}B + (I - A^{(1)}A)Y$, then with $B = AA^{(1)}B$,

$$AX = AA^{(1)}B + A(I - A^{(1)}A)Y$$

$$= B + (A - AA^{(1)}A)Y = B + 0 = B.$$

Therefore any such X is a solution.

To show that it is the general solution, we must show that if $AX_0 = B$ then $X_0 = A^{(1)}B + (I - A^{(1)}A)Y$ for some Y. Let $R = X_0 - A^{(1)}B$. Then $AR = AX_0 - AA^{(1)}B = B - B = 0$. Now therefore $R = R - A^{(1)}AR$. Hence, $X_0 = A^{(1)}B + (I - A^{(1)}A)R$ which is of the required form with $Y = R$.

In the numerical utilization of this theorem one should, of course, use some standard (1)-inverse of A such as $A^{\dagger}$.

PROBLEMS

1. Show that if A is an $m \times n$ matrix and B is any (1)-inverse of A, then AB and BA are idempotent of orders m and n respectively and BAB is a (1,2)-inverse of A.
2. Show that if A is $m \times n$ ($n \times m$), of rank m, then any (1)-inverse of A is a right (left) inverse of A, and any right (left) inverse of A is a (1,2,3)-[(1,2,4)-] inverse of A.
3. Consider two systems of equations: (1) $AX = B$, (2) $CX = D$. Find conditions such that every solution of (1) is a solution of (2).
4. What happens in Problem 3 if $B = D = 0$?
5. Prove that the matrix equation $AXB = C$ has a solution if and only if $AA^{\dagger}CB^{\dagger}B = C$. In this case, the general solution is given by

$$X = A^{\div}CB^{\div} + Y - A^{\div}AYBB^{\div}$$

for an arbitrary Y.

2.8.5 The M-P Inverse and Least Square Problems

Let A be $m \times n$, X and B be $n \times 1$, and consider the system of equations

$$AX = B.$$

If the vector B lies in the range of A, then there exists one or more solutions to this system. If the solution is not unique we might want to know which solution has minimum norm. If the vector B is not in the range of A, then there is no solution to the system, but it is often desirable to find a vector X in some way closest to a solution. To this end, for any X, define the residual vector $R = AX - B$ and consider its Euclidean norm $||R|| = \sqrt{R^*R}$. A least squares solution to the system is a vector X_0 such that its residual has minimum norm. That is,

$$||R_0|| = ||AX_0 - B|| \leq ||AX - B|| \quad \text{for all } n \times 1 \text{ vectors } X.$$

Theorem 2.8.5.1. The system of equations $AX = B$ always has a least squares solution. This solution is unique if and only if the columns of A are linearly independent. In this case, the unique least squares solution is given by $X = A^{\div}B$.

Proof. Let R(A) designate the range space of A and by $[R(A)]^{\perp}$ designate its orthogonal complement. Then we can write $B = B_1 + B_2$ where B_1 is in R(A) and B_2 is in orthogonal complement $[R(A)]^{\perp}$. For any X, AX is in R(A) as is $AX - B_1$, hence is orthogonal to B_2. Now $AX - B = AX - B_1 - B_2$. Hence, for any X, $||AX - B||^2 = ||AX - B_1||^2 + ||B_2||^2 \geq ||B_2||^2$. Therefore $||B_2||^2$ is a lower bound for the values $||AX - B||^2$ and is achieved if and only if $AX = B_1$. Since B_1 is in R(A), there is a solution X_0 to $AX = B_1$.

For this vector X_0,

$$||R_0||^2 = ||AX_0 - B||^2 = ||B_2||^2 \leq ||AX - B||^2,$$

so that the lower bound is achieved.

Since a unique solution to $AX = B_1$ exists if and only if the columns of A are linearly independent, the theorem is proved.

For any solution X_0 to $AX = B_1$,

$$R_0 = AX_0 - B = B_1 - (B_1 + B_2) = -B_2 \text{ is in } [R(A)]^{\perp}.$$

Therefore $A^*R_0 = 0$, or

$$A^*(AX_0 - B) = 0,$$

or

$$A^*AX_0 = A^*B.$$

These are the normal equations determining the least squares solution.

If the columns of A are independent, then $r(A^*A) = r(A) = n$, so that the $n \times n$ matrix A^*A is nonsingular. The least squares solution X_0 is determined by $A^*AX_0 = A^*B$, so that $X_0 = (A^*A)^{-1}A^*B$. But, from our previous work, $A^{\div} = (A^*A)^{-1}A^*$.

Finally, we take up the general case.

Lemma. Let $P = AA^{\div}$, $Q = A^{\div}A$. Then, if X and Y are arbitrary vectors (conformable),

$$||AX + (I - P)Y||^2 = ||AX||^2 + ||(I - P)Y||^2$$

and

$$||A^{\div}Y + (I - Q)X||^2 = ||A^{\div}Y||^2 + ||(I - Q)X||^2.$$

Proof. Since $A = AA^{\div}A$, $AX = AA^{\div}AX = PZ$ with $Z = AX$. We now prove that $PZ \perp (I - P)Y$. This is equivalent to $(PZ)^*(I - P)Y = 0$ or $Z^*P^*(I - P)Y = 0$. But $P^* = P$ and $P^2 = (AA^{\div}A)A^{\div} = AA^{\div} = P$. Therefore,

$P^*(I - P) = 0$. The first equality above now follows from Pythagoras' theorem. The second equality can be derived from the first using $A^{\dagger\dagger} = A$.

Another way of phrasing this work is that P is the projection onto the range space R(A) of A while I - P is the projection onto the orthogonal complement of R(A).

Theorem 2.8.5.2. Let A be m × n and B be m × 1. Let $X_0 = A^{\dagger}B$. Then for any n × 1 $X \neq X_0$, we have either

(1) $||AX - B|| > ||AX_0 - B||$

or

(2) $||AX - B|| = ||AX_0 - B||$ and $||X|| > ||X_0||$.

Proof. For any X we have

$$AX - B = AX - AA^{\dagger}B + AA^{\dagger}B - B$$
$$= A(X - A^{\dagger}B) + (I - AA^{\dagger})(-B).$$

By the previous lemma,

$$||AX - B||^2 = ||A(X - A^{\dagger}B)||^2 + ||(I - AA^{\dagger})(-B)||^2$$
$$= ||A(X - X_0)||^2 + ||AX_0 - B||^2$$
$$\geq ||AX_0 - B||^2.$$

The equality holds here if and only if $A(X - X_0) = 0$. Hence if $AX \neq AX_0$, inequality (1) holds.

Suppose, then, that $AX = AX_0$. Then $A^{\dagger}AX = A^{\dagger}AX_0 = A^{\dagger}AA^{\dagger}B = A^{\dagger}B = X_0$. Therefore, $X = X_0 + (X - X_0) = A^{\dagger}B + (I - A^{\dagger}A)X$. Hence by inequality (2) of the lemma,

$$||X||^2 = ||X_0||^2 + ||X - X_0||^2,$$

so that

$||X|| \geq ||X_0||$ and $||X|| = ||X_0||$ only if $X = X_0$.

This theorem may be rephrased as follows. Given the system AX = B. Then the vector $A^{\dagger}B$ is either the unique least squares solution or it is the least squares solution of minimum norm.

PROBLEM

1. A is square and singular. Characterize the solution $A^{\dagger}B$.

2.9 NORMAL MATRICES, QUADRATIC FORMS, AND FIELD OF VALUES

We record here a number of important facts. By a normal matrix is meant a square matrix A for which

(2.9.1) $AA^* = A^*A$.

Examples. Hermitian, skew-Hermitian, and unitary matrices are normal. Hence real symmetric, skew-symmetric, and orthogonal matrices are also normal. All circulants are normal, as we shall see.

Theorem 2.9.1. A is normal if and only if there is a unitary U and diagonal D such that $A = U^*DU$.

Theorem 2.9.2. A is normal if and only if there is a polynomial p(x) such that $A^* = p(A)$.

Theorem 2.9.3. A is Hermitian if and only if there is a unitary matrix U and a real diagonal D such that $A = U^*DU$.

Theorem 2.9.4. A is (real) symmetric if and only if there is a (real) orthogonal matrix U and a real diagonal D such that $A = U^*DU$.

PROBLEMS

1. Prove that A is normal if and only if A = R + iS where R and S are real symmetric and commute.
2. Prove that A is normal if and only if in the polar decomposition of A (A = HU with H positive semi-definite Hermitian, U unitary) one has HU = UH.
3. Let A have eigenvalues $\lambda_1, \ldots, \lambda_n$. Prove that A is normal if and only if the eigenvalues of AA* are $|\lambda_1|^2, |\lambda_2|^2, \ldots, |\lambda_n|^2$.
4. Prove that A is normal if and only if the eigenvalues of A + A* are $\lambda_1 + \overline{\lambda}_1, \lambda_2 + \overline{\lambda}_2, \ldots, \lambda_n + \overline{\lambda}_n$.
5. If A is normal and p(z) is a polynomial, then p(A) is normal.
6. If A is normal, prove that $A^{\div}$ is normal.
7. If A and B are normal, prove that A ⊗ B is normal.
8. Use Theorem 2.9.1 to prove Theorem 2.9.2.

Quadratic Forms. Let M be n × n and let $Z = (z_1, z_2, \ldots, z_n)^T$. By a quadratic form is meant the function of $z_1, \ldots, z_n$ given by

$$M(Z) = Z^*MZ. \tag{2.9.2}$$

It is often of importance to distinguish the quadratic form from a matrix that gives rise to it. The real and the complex cases are essentially different.

Lemma 2.9.5. Let Q be real and square and U a real column. Then $U^TQU = 0$ for all U if and only if $Q = -Q^T$, that is, if and only if Q is skew-symmetric.

Proof.

(a) Let $Q = -Q^T$. If $\alpha = U^TQU$, $\alpha^T = \alpha = U^TQ^TU = U^T(-Q)U = -\alpha$. Therefore $\alpha = 0$.

(b) Let $U^TQU = 0$ for all (real) U. Write $Q = Q_1 + Q_2$ where Q_1 is symmetric and Q_2 is skew-symmetric. Then, for all U

$$U^TQU = U^TQ_1U + U^TQ_2U = U^TQ_1U = 0.$$

Since Q_1 is symmetric, we have for some orthogonal P and real diagonal matrix Λ: $Q_1 = P^T\Lambda P$. Therefore for all real U, $U^TP^T\Lambda PU = (PU)^T\Lambda(PU)$. Write $PU = (\hat{u}_1, \ldots, \hat{u}_n)^T$, $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Then we have $\sum_{k=1}^{n} \lambda_k(\hat{u}_n)^2 = 0$ for all $(u_1, \ldots, u_n)$, hence for all $(\hat{u}_1, \ldots, \hat{u}_n)$. This clearly implies $\lambda_k = 0$, for k = 1, 2, ..., n. Hence $Q_1 = 0$ and $Q = Q_2$ = skew-symmetric.

Theorem 2.9.6. Let Q and R be real square and U be a real column. Then $U^TQU = U^TRU$ for all U if and only if Q - R is skew-symmetric.

Proof. $U^TQU = U^TRU$ if and only if $U^T(Q - R)U = 0$.

Corollary. Let Q be real and U be a real column. Then

$$U^TQU = U^T\left(\frac{Q + Q^T}{2}\right)U. \tag{2.9.3}$$

The matrix $\frac{1}{2}(Q + Q^T)$ is known as the symmetrization of Q.

We pass now to the complex case.

Lemma 2.9.7. Let M be a square matrix with complex elements and let Z be a column with complex elements. Then

$$Z^*MZ = 0$$

for all complex Z if and only if M = 0.

Proof

(a) The "if" is trivial.
(b) "Only if." Write Z = X + iY, M = R + iS

where X, Y, R, S are all real. Then we are given

(2.9.4) $(X^* - iY^*)(R + iS)(X + iY) = 0$ for all real X, Y.

Select Y = 0. Then $X^*(R + iS)X = 0$ for all real X or $X^*RX = 0$ and $X^*SX = 0$. Therefore, by the first lemma, R and S must be skew-symmetric: $R + R^T = 0$, $S + S^T = 0$. Expanding the product on the left side of (2.9.4), we obtain

$$X^*RX + iX^*RY + iX^*SX - X^*SY - iY^*RX + Y^*RY + Y^*SX + iY^*SY.$$

In view of the skew symmetry of R and S and the first lemma, we have $X^*RX = X^*SX = Y^*RY = Y^*SY = 0$. Therefore, we have for all real X, Y:

$$(Y^*SX - X^*SY) + i(X^*RY - Y^*RX) = 0$$

or

$$Y^*SX = X^*SY = Y^*S^*X$$

and

$$X^*RY = Y^*RX = X^*R^*Y.$$

Thus, for all real X, Y, $X^*(R - R^*)Y = 0$ and $Y^*(S - S^*)X = 0$. Selecting X and Y as appropriate unit vectors (0,.,0, 1, 0, ..., 0), this tells us that $R - R^* = 0$ and $S - S^* = 0$. But $R^* = R^T = -R$ and $S^* = S^T = -S$, therefore $R = S = 0$ and $M = 0$.

Theorem 2.9.8. Let M and N be square matrices of order n with complex elements and suppose that

(2.9.5) $Z^*MZ = Z^*NZ$

for all complex vectors Z. Then $M = N$.

Proof. As before, $Z^*MZ = Z^*NZ$ if and only if $Z^*(M - N)Z = 0$.

Note that this theorem is false if (2.9.5) holds only for real Z.

Corollary. Z^*MZ is real for all complex Z if and only if M is Hermitian.

Proof. Z^*MZ is real if and only if $Z^*MZ = (Z^*MZ)^* = Z^*M^*Z$. Hence $M = M^*$.

Let M be a Hermitian matrix. It is called positive definite if $Z^*MZ > 0$ for all $Z \neq 0$. It is called positive semidefinite if $Z^*MZ \geq 0$ for all Z. It is called indefinite if there exist $Z_1 \neq 0$ and $Z_2 \neq 0$ such that $Z_1^*MZ_1 > 0 > Z_2^*MZ_2$.

Theorem 2.9.9. Let M be a Hermitian matrix of order n with eigenvalues $\lambda_1, \ldots, \lambda_n$. Then

(a) M is positive definite if and only if $\lambda_k > 0$, $k = 1, 2, \ldots, n$.

(b) M is positive semidefinite if and only if $\lambda_k \geq 0$, $k = 1, 2, \ldots, n$.

(c) M is indefinite if and only if there are integers j, k, $j \neq k$, with $\lambda_j > 0$, $\lambda_k < 0$.

Field of Values. Let M designate a matrix of order n. The set of all complex numbers Z^*MZ with $||Z|| = 1$ is known as the field of values of M and is designated by $\mathcal{F}(M)$. $||Z||$ designates the Euclidean norm of Z.

The following facts, due to Hausdorff and Toeplitz, are known.

(1) $\mathcal{F}(M)$ is a closed, bounded, connected, convex subset of the complex plane.

(2) The field of values is invariant under unitary transformations:

(2.9.6) $\mathcal{F}(M) = \mathcal{F}(U^*MU)$, U = unitary.

(3) If ch M designates the convex hull of the eigenvalues of M, then

(2.9.7) ch M $\subseteq \mathcal{F}(M)$.

(4) If M is normal, then $\mathcal{F}(M) =$ ch M.

PROBLEMS

1. Show that the field of values of a 2 × 2 matrix M is either an ellipse (circle), a straight line segment, or a single point. More specifically, by Schur's theorem**, if one reduces M unitarily to upper triangular form,

$$M = U^*\begin{pmatrix} \lambda_1 & m \\ 0 & \lambda_2 \end{pmatrix} U, \qquad U \text{ unitary},$$

then

(a) M is not normal if and only if $m \neq 0$.

(a') $\lambda_1 \neq \lambda_2$. $\mathcal{F}(M)$ is the interior and boundary of an ellipse with foci at λ_1, λ_2, length of minor axis is $|m|$. Length of major axis $(|m|^2 + |\lambda_1 - \lambda_2|^2)^{1/2}$.

(a") $\lambda_1 = \lambda_2$. $\mathcal{F}(M)$ is the disk with center at λ_1 and radius $|m|/2$.

(b) M is normal ($m = 0$).

(b') $\lambda_1 \neq \lambda_2$. $\mathcal{F}(M)$ is the line segment joining λ_1 and λ_2.

(b") $\lambda_1 = \lambda_2$. $\mathcal{F}(M)$ is the single point λ_1.

REFERENCES

General: Aitken, [1]; Barnett and Story; Bellman, [2]; Browne; Eisele and Mason; Forsythe and Moler; Gantmacher; Lancaster, [1]; MacDuffee; Marcus; Marcus and Minc; Muir and Metzler; Newman; M. Pearl; Pullman; Suprunenko and Tyshkevich; Todd; Turnbull and Aitken.

Vandermonde matrices: Gautschi.

Discrete Fourier transforms: Aho, Hopcroft and Ullman; Carlitz; Davis and Rabinowitz; Fiduccia; Flinn and McCowan; Harmuth; Nussbaumer; Winograd; J. Pearl.

**Any square matrix is unitarily similar to an upper triangular matrix.

Hadamard matrices: Ahmed and Rao; Hall; Harmuth; Wallis, Street, and Wallis.

Generalized inverses: Ben-Israel and Greville; Meyer.

UDV theorem: Ben-Israel and Greville; Forsythe and Moler; Golub and Reinsch (numerical methods).