

# Matrix Algebra

L. R. Schaeffer, March 1999

## Definitions

1. **Scalars.** A scalar is a single quantity usually designated by an italicized lower case letter, for example  $a$ ,  $r$ , and  $k$ .
2. **Matrix.** A matrix is a group of elements arranged in rows and columns. Elements can be scalars, mathematical expressions, or other matrices. Upper case letters in bold face type denote matrices. Examples of matrices are

$$\mathbf{A} = \begin{pmatrix} 7 & 18 & -2 & 22 \\ -16 & 3 & 55 & 1 \\ 9 & -4 & 0 & 31 \end{pmatrix},$$

$$\mathbf{B} = \begin{pmatrix} x & y+1 & x+y+z \\ a-b & c \log d & e \\ \sqrt{x-y} & (m+n)/n & p \end{pmatrix},$$

and

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix}.$$

The **order** or **dimension** of a matrix is the number of rows and columns.

3. **Vectors.** A column vector is a matrix with one column, and a row vector is a matrix with one row. Usually if the word vector is used, it is implied that the vector is a column vector. Vectors are commonly denoted by lower case letters in bold face type, but this is not a universal practice.
4. **Transposition.** Transposition is the formation of a new matrix by making rows of the original matrix into columns of the new matrix in corresponding order. The transpose of a column vector is a row vector, and the transpose of a matrix of order  $m \times n$  results in a new matrix of order  $n \times m$ . If  $\mathbf{A}$  is the original matrix, then the transpose is denoted as  $\mathbf{A}'$  in these notes, but in some places the transpose is denoted as  $\mathbf{A}^T$ .
5. **Diagonals.** The diagonal elements of a matrix are those elements whose row and column numbers are the same. For example, if

$$\mathbf{D} = \begin{pmatrix} 8 & 3 & -2 \\ 5 & 7 & -1 \end{pmatrix}$$

then the diagonal elements are 8 and 7.

Off-diagonal elements of a matrix are all of the other elements excluding the diagonals.

## Types of Matrices

1. **Rectangular.** A matrix is rectangular if the number of rows does not equal the number of columns. All vectors are rectangular matrices.
2. **Square.** Square matrices have the same number of rows as columns.
3. **Symmetric.** Symmetric matrices are also square matrices. The element in the  $i^{th}$  row and  $j^{th}$  column is equal to the element in the  $j^{th}$  row and  $i^{th}$  column, for all values of  $i$  and  $j$ . The transpose of a symmetric matrix is that same matrix, i.e.  $\mathbf{A}' = \mathbf{A}$ . Take any matrix, say  $\mathbf{B}$  and form the product,  $\mathbf{B}'\mathbf{B}$ , then that product is symmetric and square.
4. **Diagonal.** A diagonal matrix is a square, symmetric matrix in which all of the elements are zero except the diagonal elements. A diagonal matrix can be written as

$$\mathbf{D} = \text{diag}(d_1 \ d_2 \ \cdots \ d_n)$$

where  $d_i$  are the diagonal elements.

5. **Identity.** An identity matrix is a special diagonal matrix in which all of the diagonal elements are equal to one. An identity matrix is usually represented by  $\mathbf{I}$ . Because it is a diagonal matrix, an identity matrix is also square and symmetric.
6. **Null.** A null matrix is a matrix of any dimension in which all elements are equal to zero, and is denoted by  $\mathbf{0}$ .
7. **Matrices With Only 1's.** A column vector of length  $n$  with every element equal to 1, is usually denoted as  $\mathbf{1}_n$ . A matrix with  $r$  rows and  $c$  columns and all elements equal to 1 is denoted as  $\mathbf{J}$ .
8. **Triangular.** Triangular matrices are square matrices with either all off-diagonal elements above or below the diagonals equal to zero. An upper triangular matrix has all elements below the diagonal equal to zero, and a lower triangular matrix has all elements above the diagonals equal to zero.
9. **Tridiagonal.** A tridiagonal matrix is a square matrix with all elements equal to zero except the diagonals and the elements immediately to the left and right of the diagonal. An example is shown below.

$$\mathbf{B} = \begin{pmatrix} 10 & 3 & 0 & 0 & 0 & 0 \\ 3 & 10 & 3 & 0 & 0 & 0 \\ 0 & 3 & 10 & 3 & 0 & 0 \\ 0 & 0 & 3 & 10 & 3 & 0 \\ 0 & 0 & 0 & 3 & 10 & 3 \\ 0 & 0 & 0 & 0 & 3 & 10 \end{pmatrix}$$

## Addition of Matrices

Matrices are *conformable for addition* if they have the same order. The resulting sum is a matrix having the same number of rows and columns as the two matrices to be added. Matrices that are not of the same order cannot be added together. If  $\mathbf{A} = \{\mathbf{a}_{ij}\}$  and  $\mathbf{B} = \{\mathbf{b}_{ij}\}$ , then

$$\mathbf{A} + \mathbf{B} = \{\mathbf{a}_{ij} + \mathbf{b}_{ij}\}.$$

For example, let

$$\mathbf{A} = \begin{pmatrix} 4 & 5 & 3 \\ 6 & 0 & 2 \end{pmatrix} \text{ and } \mathbf{B} = \begin{pmatrix} 1 & 0 & 2 \\ 3 & 4 & 1 \end{pmatrix}$$

then

$$\begin{aligned} \mathbf{A} + \mathbf{B} &= \begin{pmatrix} 4+1 & 5+0 & 3+2 \\ 6+3 & 0+4 & 2+1 \end{pmatrix} \\ &= \begin{pmatrix} 5 & 5 & 5 \\ 9 & 4 & 3 \end{pmatrix} = \mathbf{B} + \mathbf{A}. \end{aligned}$$

Subtraction is the addition of two matrices, one of which has all elements multiplied by a minus one (-1). That is,

$$\mathbf{A} + (-1)\mathbf{B} = \begin{pmatrix} 3 & 5 & 1 \\ 3 & -4 & 1 \end{pmatrix}.$$

## Multiplication of Matrices

Two matrices are *conformable for multiplication* if the number of columns in the first matrix equals the number of rows in the second matrix. If  $\mathbf{C}$  has order  $p \times q$  and  $\mathbf{D}$  has order  $m \times n$ , then the product  $\mathbf{CD}$  exists only if  $q = m$ . The product matrix has order  $p \times n$ . In general,  $\mathbf{CD}$  does not equal  $\mathbf{DC}$ , and most often the product  $\mathbf{DC}$  may not even exist because  $\mathbf{D}$  may not be conformable for multiplication with  $\mathbf{C}$ . Thus, the ordering of matrices in a product must be carefully and precisely written.

The computation of a product is defined as follows: let

$$\mathbf{C}_{p \times q} = \{c_{ij}\}$$

and

$$\mathbf{D}_{m \times n} = \{d_{ij}\}$$

and  $q = m$ , then

$$\mathbf{CD}_{p \times n} = \left\{ \sum_{k=1}^m c_{ik} d_{kj} \right\}.$$

As an example, let

$$\mathbf{C} = \begin{pmatrix} 6 & 4 & -3 \\ 3 & 9 & -7 \\ 8 & 5 & -2 \end{pmatrix} \text{ and } \mathbf{D} = \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ 3 & -1 \end{pmatrix},$$

then

$$\mathbf{CD} = \begin{pmatrix} 6(1) + 4(2) - 3(3) & 6(1) + 4(0) - 3(-1) \\ 3(1) + 9(2) - 7(3) & 3(1) + 9(0) - 7(-1) \\ 8(1) + 5(2) - 2(3) & 8(1) + 5(0) - 2(-1) \end{pmatrix} = \begin{pmatrix} 5 & 9 \\ 0 & 10 \\ 12 & 10 \end{pmatrix}.$$

## Transpose of a Product

The transpose of the product of two or more matrices is the product of the transposes of each matrix in reverse order. That is, the transpose of  $\mathbf{CDE}$ , for example, is  $\mathbf{E'D'C'}$ . The reader should verify that the result is *conformable for multiplication*, and that the result has the correct dimensions.

## Special Products

A matrix, say  $\mathbf{A}$ , is *idempotent* if the product of the matrix with itself equals itself, i.e.  $\mathbf{AA} = \mathbf{A}$ . This implies that the matrix must be square, but not necessarily symmetric.

A matrix is *nilpotent* if the product of the matrix with itself equals a null matrix, i.e.  $\mathbf{BB} = \mathbf{0}$ , then  $\mathbf{B}$  is nilpotent.

A matrix is *orthogonal* if the product of the matrix with its transpose equals an identity matrix, i.e.  $\mathbf{UU}' = \mathbf{I}$ , which also implies that  $\mathbf{U}'\mathbf{U} = \mathbf{I}$ .

### Some Rules on Multiplication

1. Multiplication of a matrix by a scalar results in multiplying every element of the matrix by that scalar.
2.  $\mathbf{ABC} = \mathbf{A(BC)} = (\mathbf{AB})\mathbf{C}$ .
3.  $\mathbf{A(B + C)} = \mathbf{AB + AC}$ .
4.  $(\mathbf{A + B})^2 = (\mathbf{A + B})(\mathbf{A + B}) = \mathbf{AA + AB + BA + BB}$ .

### Traces of Square Matrices

The trace is an operation that is applied to a square matrix and is denoted by the letters *tr*. The trace operation is the sum of the diagonal elements of a matrix. The sum is a scalar quantity. Let

$$\mathbf{A} = \begin{pmatrix} .51 & -.32 & -.19 \\ -.28 & .46 & -.14 \\ -.21 & -.16 & .33 \end{pmatrix},$$

then the trace is

$$\text{tr}(\mathbf{A}) = .51 + .46 + .33 = 1.30.$$

The trace of the product of conformable matrices has a special property known as the *rotation rule of traces*. That is,

$$\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{BCA}) = \text{tr}(\mathbf{CAB}).$$

The traces are equal, because they are scalars, even though the dimensions of the three products might be greatly different.

### Euclidean Norm

The Euclidean Norm is a matrix operation usually used to determine the degree of difference between two matrices of the same order. The norm is a scalar and is denoted as  $\|\mathbf{A}\|$ .

$$\|\mathbf{A}\| = [\text{tr}(\mathbf{AA}')]^{.5} = \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij}^2\right)^{.5}.$$

For example, let

$$\mathbf{A} = \begin{pmatrix} 7 & -5 & 2 \\ 4 & 6 & 3 \\ 1 & -1 & 8 \end{pmatrix},$$

then

$$\begin{aligned} \|\mathbf{A}\| &= (49 + 25 + 4 + \cdots + 1 + 64)^{.5} \\ &= (205)^{.5} = 14.317821. \end{aligned}$$

Other types of norms also exist.

### Other Useful Matrix Operations

1. **Direct Sum.** For matrices of any dimension, say  $\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_n$ , the direct sum is

$$\sum_i^+ \mathbf{H}_i = \mathbf{H}_1 \oplus \mathbf{H}_2 \oplus \cdots \oplus \mathbf{H}_n = \begin{pmatrix} \mathbf{H}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{H}_n \end{pmatrix}.$$

2. **Kronecker Product.** The Kronecker product, also known as the direct product, is where every element of the first matrix is multiplied, as a scalar, times the second

matrix. Suppose that  $\mathbf{B}$  is a matrix of order  $m \times n$  and that  $\mathbf{A}$  is of order  $2 \times 2$ , then the direct product of  $\mathbf{A}$  times  $\mathbf{B}$  is

$$\mathbf{A} \star \mathbf{B} = \mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} \end{pmatrix}.$$

Notice that the dimension of the example product is  $2m \times 2n$ .

3. **Hadamard Product.** The Hadamard product exists for two matrices that are conformable for addition. The corresponding elements of the two matrices are multiplied together. The order of the resulting product is the same as the matrices in the product. For two matrices,  $\mathbf{A}$  and  $\mathbf{B}$  of order  $2 \times 2$ , then the Hadamard product is

$$\mathbf{A} \odot \mathbf{B} = \begin{pmatrix} a_{11}b_{11} & a_{12}b_{12} \\ a_{21}b_{21} & a_{22}b_{22} \end{pmatrix}.$$

## Elementary Operators

Elementary operator matrices are identity matrices that have been modified by one of three methods. The product of an elementary operator matrix with any other matrix does not alter the *rank* of that matrix. Elementary operator matrices reduce a matrix to either a triangular or a diagonal form. The reduction process may involve many elementary operator matrices multiplied together. To illustrate the three types of elementary operator matrices and the process of reducing a matrix, begin with an example matrix,  $\mathbf{A}$ , where

$$\mathbf{A} = \begin{pmatrix} 4 & 16 & 24 \\ 3 & -1 & -2 \\ 1 & 5 & 8 \end{pmatrix}.$$

The first type of elementary operator matrix has one of the diagonal elements of an identity matrix replaced with a constant other than 1. Let  $\mathbf{E}_{11}(.25)$  be an elementary operator where the first diagonal element has been changed to .25. Pre-multiplication of  $\mathbf{A}$  by this matrix gives  $\mathbf{A}_1$ , where

$$\begin{aligned} \mathbf{A}_1 &= \begin{pmatrix} .25 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 4 & 16 & 24 \\ 3 & -1 & -2 \\ 1 & 5 & 8 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 4 & 6 \\ 3 & -1 & -2 \\ 1 & 5 & 8 \end{pmatrix}. \end{aligned}$$

The second type of elementary operator matrix interchanges rows or columns of a matrix. Let  $\mathbf{E}_{ij}$  be an elementary operator where rows  $i$  and  $j$  are interchanged. Pre-

multiplication of  $\mathbf{A}_1$  by this matrix gives  $\mathbf{A}_2$  where

$$\begin{aligned}\mathbf{A}_2 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \mathbf{A}_1 \\ &= \begin{pmatrix} 1 & 4 & 6 \\ 1 & 5 & 8 \\ 3 & -1 & -2 \end{pmatrix}.\end{aligned}$$

The third type of elementary operator matrix combines the elements of two rows or columns of a matrix by addition or subtraction. Let  $\mathbf{E}_{ij}(c)$  be an elementary operator where  $c$  times row  $i$  is added to row  $j$ . Pre-multiplication of  $\mathbf{A}_2$  by this matrix gives  $\mathbf{A}_3$  where

$$\begin{aligned}\mathbf{A}_3 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix} \mathbf{A}_2 \\ &= \begin{pmatrix} 1 & 4 & 6 \\ 1 & 5 & 8 \\ 0 & -13 & -20 \end{pmatrix}.\end{aligned}$$

By continued pre-multiplication of  $\mathbf{A}_3$  by other elementary operators,  $\mathbf{A}$  may be reduced to an upper triangular matrix. Further post-multiplication by another set of elementary operators can reduce  $\mathbf{A}$  to a diagonal matrix.

## Matrix Inversion

By definition, an inverse of a square matrix  $\mathbf{A}$ , denoted by  $\mathbf{A}^{-1}$ , is a matrix which when pre- or post- multiplied times the original matrix yields an identity matrix. That is,

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}, \text{ and } \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}.$$

Any square matrix with a nonzero determinant can be inverted. Thus, the first step to matrix inversion is the calculation of the determinant of the matrix. The determinant is a scalar quantity. For a  $2 \times 2$  matrix, say

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

then the determinant is

$$|\mathbf{A}| = a_{11}a_{22} - a_{21}a_{12}.$$

For a  $3 \times 3$  matrix, the determinant can be reduced to a series of determinants of  $2 \times 2$  matrices. For example, let

$$\mathbf{B} = \begin{pmatrix} 6 & -1 & 2 \\ 3 & 4 & -5 \\ 1 & 0 & -2 \end{pmatrix},$$

then

$$\begin{aligned}
 |\mathbf{B}| &= 6 \begin{vmatrix} 4 & -5 \\ 0 & -2 \end{vmatrix} - 1(-1) \begin{vmatrix} 3 & -5 \\ 1 & -2 \end{vmatrix} + 2 \begin{vmatrix} 3 & 4 \\ 1 & 0 \end{vmatrix} \\
 &= 6(-8) + 1(-1) + 2(-4) \\
 &= -57.
 \end{aligned}$$

The general expression for the determinant of a matrix is

$$|\mathbf{B}| = \sum_{j=1}^n (-1)^{i+j} b_{ij} |\mathbf{M}_{ij}|$$

where  $\mathbf{B}$  is of order  $n$ , and  $\mathbf{M}_{ij}$  is a *minor* submatrix of  $\mathbf{B}$  resulting from the deletion of the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column of  $\mathbf{B}$ . Any row of  $\mathbf{B}$  may be used to compute the determinant because the result should be the same for each row. Columns may also be used instead of rows. The reader may verify this assertion.

If the determinant is non-zero, then proceed to the next step which is to find the matrix of *signed minors*, known as the *adjoint matrix*. Using the same matrix,  $\mathbf{B}$  above, the minors and their determinants are as follows:

$$\begin{aligned}
 \mathbf{M}_{11} &= +1 \begin{pmatrix} 4 & -5 \\ 0 & -2 \end{pmatrix}, \text{ and } |\mathbf{M}_{11}| = -8, \\
 \mathbf{M}_{12} &= -1 \begin{pmatrix} 3 & -5 \\ 1 & -2 \end{pmatrix}, \text{ and } |\mathbf{M}_{12}| = +1, \\
 \mathbf{M}_{13} &= +1 \begin{pmatrix} 3 & 4 \\ 1 & 0 \end{pmatrix}, \text{ and } |\mathbf{M}_{13}| = -4, \\
 \mathbf{M}_{21} &= -1 \begin{pmatrix} -1 & 2 \\ 0 & -2 \end{pmatrix}, \text{ and } |\mathbf{M}_{21}| = -2, \\
 \mathbf{M}_{22} &= +1 \begin{pmatrix} 6 & 2 \\ 1 & -2 \end{pmatrix}, \text{ and } |\mathbf{M}_{22}| = -14, \\
 \mathbf{M}_{23} &= -1 \begin{pmatrix} 6 & -1 \\ 1 & 0 \end{pmatrix}, \text{ and } |\mathbf{M}_{23}| = -1, \\
 \mathbf{M}_{31} &= +1 \begin{pmatrix} -1 & 2 \\ 4 & -5 \end{pmatrix}, \text{ and } |\mathbf{M}_{31}| = -3, \\
 \mathbf{M}_{32} &= -1 \begin{pmatrix} 6 & 2 \\ 3 & -5 \end{pmatrix}, \text{ and } |\mathbf{M}_{32}| = 36, \text{ and} \\
 \mathbf{M}_{33} &= +1 \begin{pmatrix} 6 & -1 \\ 3 & 4 \end{pmatrix}, \text{ and } |\mathbf{M}_{33}| = 27.
 \end{aligned}$$

The adjoint matrix of signed minors,  $\mathbf{M}_B$  is then

$$\mathbf{M}_B = \begin{pmatrix} -8 & 1 & -4 \\ -2 & -14 & -1 \\ -3 & 36 & 27 \end{pmatrix}.$$

The inverse of  $\mathbf{B}$  is then

$$\begin{aligned} \mathbf{B}^{-1} &= |\mathbf{B}|^{-1} \mathbf{M}'_B \\ &= \frac{1}{-57} \begin{pmatrix} -8 & -2 & -3 \\ 1 & -14 & 36 \\ -4 & -1 & 27 \end{pmatrix}. \end{aligned}$$

If the determinant is zero, then the inverse is not defined or does not exist. A square matrix with a non-zero determinant is said to be *nonsingular*. Only nonsingular matrices have inverses. Matrices with zero determinants are called *singular* and do not have an inverse.

An important result is that the inverse of an inverse matrix is equal to the original matrix. That is,

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}.$$

The inverse of the product of two or more nonsingular matrices follows a rule similar to that for the transpose of a product of matrices. Let  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$  be three nonsingular matrices, then

$$(\mathbf{ABC})^{-1} = \mathbf{C}^{-1} \mathbf{B}^{-1} \mathbf{A}^{-1},$$

and

$$\mathbf{C}^{-1} \mathbf{B}^{-1} \mathbf{A}^{-1} \mathbf{ABC} = \mathbf{I}.$$

## Generalized Inverses

Suppose we have a set of equations,  $\mathbf{Ax} = \mathbf{r}$ , then a solution for  $\mathbf{x}$  is obtained by pre-multiplying both sides of the equation by  $\mathbf{A}^{-1}$  provided that  $\mathbf{A}$  is a nonsingular matrix. Then  $\mathbf{x} = \mathbf{A}^{-1}\mathbf{r}$ . Often, however,  $\mathbf{A}$  is a singular matrix, i.e. the determinant is zero and the inverse does not exist. Solutions to the set of equations can still be calculated using a *generalized inverse* of  $\mathbf{A}$ . A generalized inverse, denoted as  $\mathbf{A}^-$ , is a matrix that satisfies the following expression

$$\mathbf{AA}^- \mathbf{A} = \mathbf{A}.$$

In general, neither  $\mathbf{AA}^-$  nor  $\mathbf{A}^- \mathbf{A}$  are equal to  $\mathbf{I}$  except in the case when  $\mathbf{A}$  is nonsingular.

There are an infinite number of generalized inverses for any one singular matrix. A unique generalized inverse is the Moore-Penrose inverse which satisfies the following conditions:

1.  $\mathbf{AA}^- \mathbf{A} = \mathbf{A}$ ,

2.  $\mathbf{A}^{-}\mathbf{A}\mathbf{A}^{-} = \mathbf{A}^{-}$ ,
3.  $(\mathbf{A}^{-}\mathbf{A})' = \mathbf{A}^{-}\mathbf{A}$ , and
4.  $(\mathbf{A}\mathbf{A}^{-})' = \mathbf{A}\mathbf{A}^{-}$ .

Usually, a generalized inverse that satisfies only the first condition is sufficient for practical purposes.

- **Linear Independence.** A square matrix with a nonzero determinant is a matrix in which the rows and columns of the matrix are linearly independent. If  $\mathbf{A}$  is the matrix, then linear independence means that no vector, say  $\mathbf{k}$ , exists such that  $\mathbf{A}\mathbf{k} = \mathbf{0}$  except for  $\mathbf{k} = \mathbf{0}$ .

A square matrix with a zero determinant implies that at least one non-null vector,  $\mathbf{k}$ , exists such that  $\mathbf{A}\mathbf{k} = \mathbf{0}$ . To compute a generalized inverse of a matrix that has a zero determinant requires that we know the number of linearly independent rows and columns in that matrix. The number of linearly independent rows or columns is called the *rank* of a matrix and is a term that applies to rectangular as well as square matrices.

- **Rank.** Elementary operators are used to determine the rank of a matrix. Reduction of the matrix to a diagonal matrix is called *reduction to canonical form*, and the reduced matrix is called the *canonical form under equivalence*. If the matrix is square and symmetric, then the reduction of the matrix to a diagonal form with either ones or minus ones on the diagonals is called *congruent reduction* and the resulting matrix is the *canonical form under congruence*. Usually, only a reduction to an upper triangular form is necessary to determine the rank of the matrix.

Let

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & -1 & 3 & 8 \\ 4 & 7 & 5 & -2 & -1 \\ 6 & 8 & 4 & 1 & 7 \end{pmatrix}.$$

The rank of a matrix can not be greater than the minimum of the number of rows or columns, whichever is smaller. In the example above,  $\mathbf{A}$  has 3 rows and 5 columns, and therefore the rank of  $\mathbf{A}$  can not be greater than 3. To find the exact rank, elementary operator matrices are used. Let  $\mathbf{P}_1$  be the first elementary operator matrix that will zero out the first elements in the second and third rows. Premultiplication of any matrix by elementary operator matrices does not change the rank of that matrix. Let

$$\mathbf{P}_1 = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix},$$

then

$$\mathbf{P}_1\mathbf{A} = \begin{pmatrix} 2 & 1 & -1 & 3 & 8 \\ 0 & 5 & 7 & -8 & -17 \\ 0 & 5 & 7 & -8 & -17 \end{pmatrix}.$$

Now use  $\mathbf{P}_2$  to subtract the third row from the second row, so that

$$\mathbf{P}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix},$$

and

$$\mathbf{P}_2\mathbf{P}_1\mathbf{A} = \begin{pmatrix} 2 & 1 & -1 & 3 & 8 \\ 0 & 5 & 7 & -8 & -17 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The rank is the number of non-zero diagonal elements of the reduced matrix, i.e.  $r(\mathbf{A}) = 2$ .

Elementary operators of order  $5 \times 5$  would be postmultiplied times  $\mathbf{A}$  in order to reduce the matrix to a diagonal form, and the rank would still be the number of non-zero diagonal elements.

- **Definitions.**

**Full-row rank** If  $\mathbf{A}$  has order  $m \times n$  with rank equal to  $m$ , then  $\mathbf{A}$  has full row rank.

**Full-column rank** A matrix with rank equal to the number of columns has full-column rank.

**Full rank** A square matrix with rank equal to the number of rows or columns has full rank. A full rank matrix is nonsingular, has a non-zero determinant, and has an inverse.

1. A null matrix has a rank of zero.
2. A **J**-matrix has a rank of one.
3. The rank of an idempotent matrix is equal to its trace.
4. If  $\mathbf{A}$  has rank  $r$  and  $\mathbf{B}$  has rank  $q$ , and the two matrices are conformable for multiplication, then the product,  $\mathbf{AB}$ , has a maximum possible rank equal to the lesser of  $r$  or  $q$ .

- **Consequences.** The fact that a matrix has rank that is less than the number of rows or columns in the matrix means that we can partition the matrix into a square matrix of order equal to the rank and that has rank equal to that matrix, and into other matrices of the remaining rows and columns. The other rows and columns are linearly dependent upon the rows and columns in that square matrix. That means that they can be formed from the rows and columns that are linearly independent.

Let  $\mathbf{A}$  be a matrix of order  $p \times q$  with rank  $r$ , and  $r$  is less than either  $p$  or  $q$ , then there are  $p - r$  rows of  $\mathbf{A}$  and  $q - r$  columns which are not linearly independent. Partition  $\mathbf{A}$  as follows:

$$\mathbf{A}_{p \times q} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$$

such that  $\mathbf{A}_{11}$  has order  $r \times r$  and rank of  $r$ . Re-arrangement of rows and columns of  $\mathbf{A}$  may be needed to find an appropriate  $\mathbf{A}_{11}$ .  $\mathbf{A}_{12}$  has order  $r \times (q - r)$ ,  $\mathbf{A}_{21}$  has order  $(p - r) \times r$ , and  $\mathbf{A}_{22}$  has order  $(p - r) \times (q - r)$ .

There exist matrices,  $\mathbf{K}_1$  and  $\mathbf{K}_2$  such that

$$\begin{pmatrix} \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} = \mathbf{K}_2 \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \end{pmatrix},$$

and

$$\begin{pmatrix} \mathbf{A}_{12} \\ \mathbf{A}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11} \\ \mathbf{A}_{21} \end{pmatrix} \mathbf{K}_1,$$

and

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{11}\mathbf{K}_1 \\ \mathbf{K}_2\mathbf{A}_{11} & \mathbf{K}_2\mathbf{A}_{11}\mathbf{K}_1 \end{pmatrix}.$$

To illustrate, let

$$\mathbf{A} = \begin{pmatrix} 3 & -1 & 4 \\ 1 & 2 & -1 \\ -5 & 4 & -9 \\ 4 & 1 & 3 \end{pmatrix},$$

and the rank of this matrix is 2. A  $2 \times 2$  full rank submatrix within  $\mathbf{A}$  is the upper left  $2 \times 2$  matrix. Let

$$\mathbf{A}_{11} = \begin{pmatrix} 3 & -1 \\ 1 & 2 \end{pmatrix},$$

then

$$\mathbf{A}_{12} = \begin{pmatrix} 4 \\ -1 \end{pmatrix} = \mathbf{A}_{11}\mathbf{K}_1,$$

where

$$\mathbf{K}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

$$\mathbf{A}_{21} = \begin{pmatrix} -5 & 4 \\ 4 & 1 \end{pmatrix} = \mathbf{K}_2\mathbf{A}_{11},$$

where

$$\mathbf{K}_2 = \begin{pmatrix} -2 & 1 \\ 1 & 1 \end{pmatrix},$$

and

$$\mathbf{A}_{22} = \mathbf{K}_2\mathbf{A}_{11}\mathbf{K}_1$$

$$= \begin{pmatrix} -2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -9 \\ 3 \end{pmatrix}.$$

This kind of partitioning of matrices with less than full rank is **always** possible. In practice, we need only know that we can perform this kind of partitioning, but it is usually not necessary to derive  $\mathbf{K}_1$  and  $\mathbf{K}_2$  explicitly.

- **Calculations.** For a matrix,  $\mathbf{A}$ , of less than full rank, there is an infinite number of possible generalized inverses, all of which would satisfy  $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$ . However, only one generalized inverse needs to be computed in practice. A method to derive one particular type of generalized inverse has the following steps:

1. Determine the rank of  $\mathbf{A}$ .
2. Obtain  $\mathbf{A}_{11}$ , a square, full rank submatrix of  $\mathbf{A}$  with rank equal to the rank of  $\mathbf{A}$ .
3. Partition  $\mathbf{A}$  as

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}.$$

4. Compute the generalized inverse as

$$\mathbf{A}^- = \begin{pmatrix} \mathbf{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

If  $\mathbf{A}$  has order  $p \times q$ , then  $\mathbf{A}^-$  must have order  $q \times p$ . To prove that  $\mathbf{A}^-$  is a generalized inverse of  $\mathbf{A}$ , then multiply out the expression

$$\begin{aligned} \mathbf{A}\mathbf{A}^-\mathbf{A} &= \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{pmatrix}. \end{aligned}$$

From the previous section,  $\mathbf{A}_{21} = \mathbf{K}_2\mathbf{A}_{11}$  so that

$$\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} = \mathbf{K}_2\mathbf{A}_{11}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} = \mathbf{K}_2\mathbf{A}_{12} = \mathbf{A}_{22}.$$

- **Solutions to Equations.** Because there are an infinite number of generalized inverses to a matrix that has less than full rank, then it logically follows that for a system of consistent equations,  $\mathbf{A}\mathbf{x} = \mathbf{r}$ , where the solutions are computed as  $\mathbf{x} = \mathbf{A}^-\mathbf{r}$ , then there would also be an infinite number of solution vectors for  $\mathbf{x}$ . Having computed only one generalized inverse, however, it is possible to compute many different solution vectors. From Searle(1971 page 9, Theorem 2), if  $\mathbf{A}$  has  $q$  columns and if  $\mathbf{G}$  is one generalized inverse of  $\mathbf{A}$ , then the consistent equations  $\mathbf{A}\mathbf{x} = \mathbf{r}$  have solution

$$\tilde{\mathbf{x}} = \mathbf{G}\mathbf{r} + (\mathbf{G}\mathbf{A} - \mathbf{I})\mathbf{z},$$

where  $\mathbf{z}$  is any arbitrary vector of length  $q$ . The number of linearly independent solution vectors, however, is  $(q - r + 1)$ .

It is also possible to generate other generalized inverses of the same matrix. If  $\mathbf{G}$  is a generalized inverse of  $\mathbf{A}$  then so is

$$\mathbf{F} = \mathbf{GAG} + (\mathbf{I} - \mathbf{GA})\mathbf{X} + \mathbf{Y}(\mathbf{I} - \mathbf{AG})$$

for any  $\mathbf{X}$  and  $\mathbf{Y}$ . Pre- and post- multiplication of  $\mathbf{F}$  by  $\mathbf{A}$  shows that this is so.

- **Special Generalized Inverses.** In linear models analyses, the product  $\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}$  occurs frequently where  $\mathbf{X}$  is a matrix that usually has more rows than it has columns and has rank less than the number of columns, and  $\mathbf{R}$  is a square matrix that is usually diagonal. Generalized inverses of this product matrix have special features that are used in the derivation and proof of various statistical procedures. Let  $\mathbf{X}$  be a matrix of order  $N \times p$  with rank  $r$ . The product,  $\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}$  has order  $p \times p$  and is symmetric with rank  $r$ . Let  $\mathbf{G}$  represent any generalized inverse of the product matrix, then the following results are true.

1.  $\mathbf{G}$  is not necessarily a symmetric matrix, in which case  $\mathbf{G}'$  is also a generalized inverse of the product matrix.
2.  $\mathbf{X}'\mathbf{R}^{-1}\mathbf{XGX}' = \mathbf{X}'$  or  $\mathbf{XGX}'\mathbf{R}^{-1}\mathbf{X} = \mathbf{X}$ .
3.  $\mathbf{GX}'\mathbf{R}^{-1}$  is a generalized inverse of  $\mathbf{X}$ .
4.  $\mathbf{XGX}'$  is always symmetric and unique for all generalized inverses of the product matrix,  $\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}$ .
5. If  $\mathbf{1}' = \mathbf{k}'\mathbf{X}$  for some  $\mathbf{k}'$ , then  $\mathbf{1}'\mathbf{R}^{-1}\mathbf{XGX}' = \mathbf{1}'$ .

The proofs of the above results can be found in Searle (1971). These results are **very important**.

## Eigenvalues and Eigenvectors

There are a number of square, and sometimes symmetric, matrices involved in statistical procedures that must be *positive definite*. Suppose that  $\mathbf{Q}$  is any square matrix then

- $\mathbf{Q}$  is positive definite if  $\mathbf{y}'\mathbf{Q}\mathbf{y}$  is always greater than zero for all vectors,  $\mathbf{y}$ .
- $\mathbf{Q}$  is *positive semi-definite* if  $\mathbf{y}'\mathbf{Q}\mathbf{y}$  is greater than or equal to zero for all vectors  $\mathbf{y}$ , and for at least one vector  $\mathbf{y}$ , then  $\mathbf{y}'\mathbf{Q}\mathbf{y} = 0$ .
- $\mathbf{Q}$  is *non-negative definite* if  $\mathbf{Q}$  is either positive definite or positive semi-definite.

To determine the appropriate condition of the matrix  $\mathbf{Q}$ , then one must calculate the eigenvalues (or latent roots or characteristic roots) of the matrix. The eigenvalues are useful in that

- The product of the eigenvalues equals the determinant of the matrix.
- The number of non-zero eigenvalues equals the rank of the matrix.
- If all the eigenvalues are greater than zero, then the matrix is positive definite.
- If all the eigenvalues are greater than or equal to zero and one or more are equal to zero, then the matrix is positive semi-definite.

If  $\mathbf{Q}$  is a square, symmetric matrix, then it can be represented as

$$\mathbf{Q} = \mathbf{U}\mathbf{D}\mathbf{U}'$$

where  $\mathbf{D}$  is a diagonal matrix, the canonical form of  $\mathbf{Q}$ , containing the eigenvalues of  $\mathbf{Q}$ , and  $\mathbf{U}$  is an orthogonal matrix of the corresponding eigenvectors. Recall that for a matrix to be orthogonal then  $\mathbf{U}\mathbf{U}' = \mathbf{I} = \mathbf{U}'\mathbf{U}$ , and  $\mathbf{U}^{-1} = \mathbf{U}'$ .

The eigenvalues and eigenvectors are found by solving

$$|\mathbf{Q} - d\mathbf{I}| = 0,$$

and

$$\mathbf{Q}\mathbf{u} - d\mathbf{u} = \mathbf{0},$$

where  $d$  is one of the eigenvalues of  $\mathbf{Q}$  and  $\mathbf{u}$  is the corresponding eigenvector. There are numerous computer routines for calculating  $\mathbf{D}$  and  $\mathbf{U}$ .

## Differentiation

The differentiation of mathematical expressions involving matrices follows similar rules as for those involving scalars. Some of the basic results are shown below.

Let

$$\begin{aligned} c &= 3x_1 + 5x_2 + 9x_3 = \mathbf{b}'\mathbf{x} \\ &= \begin{pmatrix} 3 & 5 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}. \end{aligned}$$

With scalars we have

$$\begin{aligned} \frac{\partial c}{\partial x_1} &= 3 \\ \frac{\partial c}{\partial x_2} &= 5 \\ \frac{\partial c}{\partial x_3} &= 9, \end{aligned}$$

but with vectors we have

$$\frac{\partial c}{\partial \mathbf{x}} = \mathbf{b}.$$

The general rule is

$$\frac{\partial \mathbf{A}'\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}.$$

Another function might be

$$c = 9x_1^2 + 6x_1x_2 + 4x_2^2$$

or

$$c = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 9 & 3 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{x}'\mathbf{A}\mathbf{x}.$$

With scalars we would have

$$\frac{\partial c}{\partial x_1} = 2(9)x_1 + 6x_2$$

$$\frac{\partial c}{\partial x_2} = 6x_1 + 2(4)x_2,$$

then in matrix form,

$$\frac{\partial c}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}.$$

If  $\mathbf{A}$  was not a symmetric matrix, then

$$\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{A}'\mathbf{x}.$$

## Cholesky Decomposition

In simulation studies or applications of Gibb's sampling there is frequently a need to factor a symmetric positive definite matrix into the product of a matrix times its transpose. The Cholesky decomposition of a matrix, say  $\mathbf{V}$ , is a lower triangular matrix such that

$$\mathbf{V} = \mathbf{T}\mathbf{T}',$$

and  $\mathbf{T}$  is lower triangular. Suppose that

$$\mathbf{V} = \begin{pmatrix} 9 & 3 & -6 \\ 3 & 5 & 0 \\ -6 & 0 & 21 \end{pmatrix}.$$

The problem is to derive

$$\mathbf{T} = \begin{pmatrix} t_{11} & 0 & 0 \\ t_{21} & t_{22} & 0 \\ t_{31} & t_{32} & t_{33} \end{pmatrix},$$

such that

$$\begin{aligned}
 t_{11}^2 &= 9 \\
 t_{11}t_{21} &= 3 \\
 t_{11}t_{31} &= -6 \\
 t_{21}^2 + t_{22}^2 &= 5 \\
 t_{21}t_{31} + t_{22}t_{32} &= 0 \text{ and} \\
 t_{31}^2 + t_{32}^2 + t_{33}^2 &= 21
 \end{aligned}$$

These equations give  $t_{11} = 3$ , then  $t_{21} = 3/t_{11} = 1$ , and  $t_{31} = -6/t_{11} = -2$ . From the fourth equation,  $t_{22}$  is the square root of  $(5 - t_{21}^2)$  or  $t_{22} = 2$ . The fifth equation says that  $(1)(-2) + (2)t_{32} = 0$  or  $t_{32} = 1$ . The last equation says that  $t_{33}^2$  is  $21 - (-2)^2 - (1)^2 = 16$ . The end result is

$$\mathbf{T} = \begin{pmatrix} 3 & 0 & 0 \\ 1 & 2 & 0 \\ -2 & 1 & 4 \end{pmatrix}.$$

The derivation of the Cholesky decomposition is easily programmed for a computer. Note that in calculating the diagonals of  $\mathbf{T}$  the square root of a number is needed, and consequently this number must always be positive. Hence, if the matrix is positive definite, then all necessary square roots will be of positive numbers. However, the opposite is not true. That is, if all of the square roots are of positive numbers, the matrix is not necessarily guaranteed to be positive definite. The only way to guarantee positive definiteness is to calculate the eigenvalues of a matrix and to see that they are all positive.

### Inverse of a Lower Triangular Matrix

The inverse of a lower triangular matrix is also a lower triangular matrix, and can be easily derived. The diagonals of the inverse are the inverse of the diagonals of the original matrix. Using the matrix  $\mathbf{T}$  from the previous section, then

$$\mathbf{T}^{-1} = \begin{pmatrix} t^{11} & 0 & 0 \\ t^{21} & t^{22} & 0 \\ t^{31} & t^{32} & t^{33} \end{pmatrix},$$

where

$$\begin{aligned}
 t^{ii} &= 1/t_{ii} \\
 t_{21}t^{11} + t_{22}t^{21} &= 0 \\
 t_{31}t^{11} + t_{32}t^{21} + t_{33}t^{31} &= 0 \\
 t_{32}t^{22} + t_{33}t^{32} &= 0.
 \end{aligned}$$

These equations give

$$t^{11} = \frac{1}{3}$$

$$\begin{aligned}t^{21} &= -\frac{1}{6} \\t^{22} &= \frac{1}{2} \\t^{31} &= \frac{5}{24} \\t^{32} &= -\frac{1}{8} \text{ and} \\t^{33} &= \frac{1}{4}\end{aligned}$$

Likewise the determinant of a triangular matrix is the product of all of the diagonal elements. Hence all diagonal elements need to be non-zero for the inverse to exist.

The natural logarithm of the determinant of a triangular matrix is the summation of the natural logarithms of the individual diagonal elements. This property is useful in derivative free restricted maximum likelihood.