MATRIX ALGEBRA

S. R. Searle

Biometrics Unit, Cornell University, Ithaca, N. Y. 14853

BU-1355-M July 1996

Abstract

This is a 6,000–10,000 word article invited for an upcoming *Encyclopedia of Biostatistics*. It attempts to describe and summarize some of the main features of matrix algebra as are customarily found in the statistical literature.

Key Words

Basic matrix operations, special matrices, determinants, inverses, rank, canonical forms, generalized inverses, linear equations, eigenroots, eigenvectors, differential calculus, vec, vech, real and complex matrices, statistical uses.

MATRIX ALGEBRA

The algebra we learn when teenagers has letters of the alphabet each representing a number. For example: a father and son are x and y years old, respectively, and their total age is 70. In ten year's time the father will be twice as old as the son. Hence $x + y = 70$ and $x + 10 = 2(y + 10)$ and so $x = 50$ and $y = 20$.

In contrast, matrix algebra is the algebra of letters each representing many numbers, with those numbers always arrayed in the form of a rectangle (or square). An example is

$$\mathbf{X} = \begin{bmatrix} 9 & 0 & 7 & t \\ u^2 + v & -3 & 6.1 & 5^3 \end{bmatrix}.$$

GENERAL DESCRIPTION

A *matrix* is a rectangular array of numbers, which can be any mixture of numbers that are complex, real, zero, positive, negative, decimal, fractions, or algebraic expressions. When none of them is complex $\left(\text{i.e., involving } \sqrt{-1}\right)$, the matrix is said to be *real.* And because statistics deals with data, which are real numbers (especially biological data), almost all of this article applies to real matrices. Each number in a matrix is called an *element*: in being some representation of a single number it is called a *scalar*, to contrast with matrix which represents many numbers.

Elements are always set out in rows and columns with the number of rows and columns being called the *order* (or *dimension*) of the matrix. Thus the illustrated **X** has order 2×4 ("two by four") with the number of rows being mentioned first. Sometimes the order is used as a subscript to the matrix symbol, e.g., $\mathbf{X}_{2\times 4}$. In this encyclopedia the widespread custom is used of denoting matrices by bold face, capital, Roman letters.

Elements of a matrix can be represented by letters, having subscripts to denote location (row and column) in the matrix. Thus a matrix **A** might be represented as

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

The first subscript indicates row, and the second column; e.g., a_{23} is in row 2 and column 3. More briefly, we can write

$$\mathbf{A} = \{a_{ij}\} \quad \text{for} \quad i = 1, 2, 3 \quad \text{and} \quad j = 1, 2, 3 \,.$$

When **B** has r rows and c columns

$$\mathbf{B} = \{b_{ij}\} \quad \text{for} \quad i = 1, 2, \cdots, r \quad \text{and} \quad j = 1, 2, \cdots, c \,.$$

A more compact form is

$$\mathbf{B}\left\{_m \; b_{ij}\right\}_{i=1,\, j=1}^{r \quad c} ,$$

the m indicating it is a matrix. The element in the first row and first column (e.g., a_{11} in **A** and the 9 in **X**) is called the *leading element.*

By virtue of a matrix being a rectangular array there are many special forms, the first two of which are square matrices, and vectors.

Square matrices

1. Have the same number of rows as columns. **A** is an example.

2. Elements on the geometric diagonal from upper left to lower right, those with both subscripts the same, are *diagonal elements*; they constitute *the diagonal of the matrix.*

3. Elements immediately below the diagonal constitute the *sub-diagonal.*

4. Elements not on the diagonal are *off-diagonal elements.*

5. When all off-diagonal elements are zero, and at least some diagonal elements are non-zero, the matrix is a *diagonal matrix.*

6. When all elements below (above) the diagonal are zero the matrix is said to be *upper* (*lower*) *triangular.*

Vectors

When a matrix has only one column it is a *column vector* or, more usually, just *vector*; and it shall here be denoted by a bold face, lower case, Roman letter from near the end of the alphabet, e.g.,

$$\mathbf{x} = \begin{bmatrix} 1 \\ 7 \\ -4 \\ 0 \end{bmatrix},$$

When a matrix has only one row it is called a *row vector*. Notation is similar to that for a column vector, except for a superscript prime:

$$\mathbf{y}' = [0 \quad -4 \quad 9 \quad 12 \quad 37]\,.$$

BASIC OPERATIONS

A minimal requirement for matrix algebra is to define the arithmetic operations. Moreover, the rectangular nature of matrices begets numerous operations that do not exist for scalars; e.g., changing rows into columns, and columns into rows.

The transpose of a matrix

Changing $\mathbf{A}$ so that its rows become columns (and hence its columns become rows) gives a matrix called the *transpose* of $\mathbf{A}$, written traditionally as $\mathbf{A}'$ (and sometimes to-day as $\mathbf{A}^{\mathrm{T}}$). Thus for

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 6 & 1 & -2 & 5 \end{bmatrix}, \qquad \mathbf{A}' = \begin{bmatrix} 1 & 6 \\ 2 & 1 \\ 3 & -2 \\ 4 & 5 \end{bmatrix}.$$

Note that the transpose of $\mathbf{A}'$ is $\mathbf{A}$: $(\mathbf{A}')' = \mathbf{A}$. And the transpose of a column vector is a row vector (and vice versa): $[1 \quad 2 \quad 3]' = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

Partitioned matrices

The rows and columns of a matrix can be partitioned into a representation that is a matrix of matrices of smaller orders:

$$\mathbf{K} = \left[\begin{array}{cc|cc} 1 & 2 & 3 & 4 \\ 6 & 8 & 4 & 0 \\ 9 & 8 & 1 & 2 \\ \hline 6 & 8 & 3 & 9 \\ 4 & 1 & 6 & 1 \end{array}\right] = \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{bmatrix}, \quad \text{for} \quad \mathbf{K}_{11} = \begin{bmatrix} 1 & 2 \\ 6 & 8 \\ 9 & 8 \end{bmatrix}, \quad \text{and so on}\,.$$

$\mathbf{K}$ is a *partitioned* matrix; the $\mathbf{K}$s with subscripts are *sub-matrices* of $\mathbf{K}$.

In transposing a partitioned matrix, not only is the matrix of submatrices transposed but each submatrix is also transposed. Thus

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}' = \begin{bmatrix} \mathbf{A}' & \mathbf{C}' \\ \mathbf{B}' & \mathbf{D}' \end{bmatrix}.$$

A matrix can also be partitioned into its columns (or its rows): e.g.,

$$\mathbf{K} = [\mathbf{k}_1 \;\; \mathbf{k}_2 \;\; \mathbf{k}_3 \;\; \mathbf{k}_4] ,$$

where each of the subscripted **k**s is a column of **K**.

The trace of a matrix

The trace of a matrix is defined only for a square matrix; and *trace* of **A** is the sum of the diagonal elements of **A** often written as tr(**A**). Note that tr(**A**) = tr(**A**′), and tr(scalar) = scalar.

Addition and subtraction

Addition and subtraction are defined only for matrices of the same order, whereupon the matrices are said to be *conformable* for addition and subtraction. Then, for $\mathbf{A} = \{a_{ij}\}$ and $\mathbf{B} = \{b_{ij}\}$,

$$\mathbf{A} + \mathbf{B} = \{a_{ij} + b_{ij}\} .$$

If two matrices do not have the same order their sum and differences do not exist. Note the properties

$$(\mathbf{A} \pm \mathbf{B})' = \mathbf{A}' \pm \mathbf{B}' \quad \text{and} \quad \text{tr}(\mathbf{A} \pm \mathbf{B}) = \text{tr}(\mathbf{A}) \pm \text{tr}(\mathbf{B}) .$$

Scalar multiplication

For λ being a scalar, $\lambda\mathbf{A}$ is **A** with every element multiplied by λ. Thus for $\mathbf{A} = \{a_{ij}\}$, $\lambda\mathbf{A} = \{\lambda a_{ij}\}$.

Equality and null matrices

Two matrices are equal only when they are equal element by element. Thus for

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 6 & 8 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 6 & 8 \end{bmatrix} \qquad \text{and} \qquad \mathbf{C} = \begin{bmatrix} 1 & 2 \\ 5 & 8 \end{bmatrix}$$

$\mathbf{A} = \mathbf{B}$, but $\mathbf{A} \neq \mathbf{C}$. And

$$\mathbf{A} - \mathbf{B} = \begin{bmatrix} 1-1 & 2-2 \\ 6-6 & 8-8 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \mathbf{0} .$$

Any matrix having every element zero is a *null matrix*. It is a zero of matrix algebra: note, it is *a* zero not *the* zero, because null matrices can be of any order.

Multiplication

Multiplication of matrices differs greatly from that of scalars. First of all **AB** and **BA** can, and often do, differ. To distinguish between the two, **AB** is described as **B** *pre-multiplied* by **A** (or as **B** *post-multiplied* as **A**.

The *inner product* of two vectors is a row vector post-multiplied by a column vector, with both vectors having the same number of elements; for example, by definition

$$[1 \quad 7 \quad 2]\begin{bmatrix} 3 \\ 5 \\ 9 \end{bmatrix} = 1(3) + 7(5) + 2(9) = 56 \ .$$

Thus for $\mathbf{x}' = \{x_i\}_{i=1}^{n}$ and $\mathbf{y}' = \{y_i\}_{i=1}^{n}$,

$$\mathbf{x}'\mathbf{y} = \sum_{i=i}^{n} x_i y_i \ .$$

In contrast, an *outer product* is a column vector post-multiplied by a row vector

$$\mathbf{x}\mathbf{y}' = \{x_i y_j\} \ .$$

In this case the vectors can be of different orders.

The product **AB** exists only when **A** has as many column as **B** has rows. And then **A** and **B** are described as being *conformable for the product* **AB**, whereupon

$$\mathbf{A}_{r \times c}\mathbf{B}_{c \times t} = \mathbf{P}_{r \times t} \ .$$

In **P**, the element in row i and column j is the inner product of row i of **A** and column j of **B**:

$$\mathbf{P}_{r \times t} = \{p_{ij}\} = \left\{\sum_{k=1}^{c} a_{ik}b_{kj}\right\} \quad \text{for} \quad i = 1, \cdots, r \quad \text{and} \quad j = 1, \cdots, t \ .$$

Important consequences of this are that **AB** exists only for $\mathbf{A}_{r \times c}$ and $\mathbf{B}_{c \times t}$; both **AB** and **BA** exist only for $\mathbf{A}_{r \times c}$ and $\mathbf{B}_{c \times r}$, but they will be of different orders (and so not equal) unless r = c. And even then **AB** and **BA** are not necessarily equal. For example

$$\begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} 3 & 4 \\ -5 & 6 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 11 & 2 \end{bmatrix} \quad \text{but} \quad \begin{bmatrix} 3 & 4 \\ -5 & 6 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix} = \begin{bmatrix} 11 & -4 \\ 7 & -6 \end{bmatrix} .$$

Products with null matrices

Every product of a matrix with a null matrix is a null matrix: but those null matrices are not necessarily of the same order. Thus $\mathbf{0}_{3 \times 2}\mathbf{A}_{2 \times 5} = \mathbf{0}_{3 \times 5}$ and $\mathbf{A}_{2 \times 5}\mathbf{0}_{5 \times 6} = \mathbf{0}_{2 \times 6}$.

Products with diagonal matrices

Pre- (post-) multiplying **A** by a diagonal matrix **D** multiplies each row (column) of **A** by the corresponding diagonal element of **D**.

Identity matrices

If every diagonal element of a diagonal matrix is a one the matrix is called an *identity* matrix, **I**; pre- or post-multiplication of **A** by an identity matrix yields **A**. Thus **I**-matrices are the unities of matrix algebra.

Transposing a product

The transpose of a product is the product of the transposed matrices in reverse order. Thus

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}', \quad \text{and} \quad (\mathbf{XAY})' = \mathbf{Y}'\mathbf{A}'\mathbf{X}' .$$

Trace of a product

The trace of a product equals the trace of cyclic permutations of that product: $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$ and $\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{CAB}) = \text{tr}(\mathbf{BCA})$, but the latter does not equal the trace of **ACB**.

Powers of matrices

Only square matrices have powers: $\mathbf{A}_{2\times 4}\mathbf{A}_{2\times 4}$ does not exist. $\mathbf{A}_{4\times 4}\mathbf{A}_{4\times 4}$ written as $\mathbf{A}^2_{4\times 4}$ does.

Hadamard products

The (i,j)'th element of **AB** is $\sum_k a_{ik}b_{kj}$. But there are other ways of defining a product. One is the *Hadamard product* defined as

$$\mathbf{A} \cdot \mathbf{B} = \{a_{ij}b_{ij}\} .$$

Thus the (i,j)th element of the Hadamard product is the product of the (i,j)th element of **A** and **B**— which must have the same order.

Direct products

There is also the direct product

$$\mathbf{A} \otimes \mathbf{B} = \{a_{ij}\mathbf{B}\} .$$

When **A** has order $p \times q$ and **B** has order $r \times s$, $\mathbf{A} \otimes \mathbf{B}$ has order $pr \times qs$.

Laws of algebra

Providing conformability requirements are met, it is only the commutative law of algebra which is not met; i.e., **AB** equaling **BA** is far from true. Otherwise

$$(\mathbf{A}+\mathbf{B})+\mathbf{C}=\mathbf{A}+\mathbf{B}+\mathbf{C}$$

$$(\mathbf{AB})\mathbf{C}=\mathbf{A}(\mathbf{BC})=\mathbf{ABC}$$

$$\mathbf{A}(\mathbf{B}+\mathbf{C})=\mathbf{AB}+\mathbf{AC}$$

$$\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}\,.$$

Contrasts with scalar algebra

The following results illustrate differences in the algebra of matrices compared to that of scalars.

$\mathbf{AX}+\mathbf{BX}=(\mathbf{A}+\mathbf{B})\mathbf{X} \neq \mathbf{X}(\mathbf{A}+\mathbf{B})$.

XP + **QX** does *not* have **X** as a factor.

$\mathbf{AB}=\mathbf{0}$ does not imply that **A** or **B** are **0**, nor does it imply that **BA** is **0**.

$\mathbf{Y}^2=\mathbf{0}$ defines **Y** as *nilpotent* and does *not* imply that **Y** is **0**.

$\mathbf{Z}^2=\mathbf{I}$ does *not* imply that **Z** is $\pm\mathbf{I}$.

$\mathbf{Q}^2=\mathbf{Q}$ defines **Q** as *idempotent* but does not imply that **Q** is **0** or **I**.

Examples of these last four features are

$$\mathbf{AB}=\begin{bmatrix}1&1\\1&1\end{bmatrix}\begin{bmatrix}1&1\\-1&-1\end{bmatrix}=\mathbf{0}\,,\qquad \mathbf{BA}=\begin{bmatrix}1&1\\-1&-1\end{bmatrix}\begin{bmatrix}1&1\\1&1\end{bmatrix}=\begin{bmatrix}2&2\\-2&-2\end{bmatrix}$$

$$\mathbf{Y}^2=\begin{bmatrix}1&1\\-1&-1\end{bmatrix}^2=\mathbf{0}\,,\qquad \mathbf{Z}^2=\begin{bmatrix}1&0\\4&-1\end{bmatrix}^2=\mathbf{I}$$

and

$$\mathbf{Q}^2=\begin{bmatrix}3&-2\\3&-2\end{bmatrix}^2=\mathbf{Q}\,.$$

One may be tempted to think of these examples as pathological cases. To some extent they are, born of the need to have illustrations that occupy minimum space; but they serve as stern warning that what can be done in scalar algebra does not always carry over to matrix algebra.

SPECIAL MATRICES

Square matrices and vectors have already been mentioned as special forms of matrices. There are many others, some arising from their intrinsic properties, others from the applications in which they arose. Just a few of the more commonly occurring ones are mentioned here.

Symmetric matrices

A is defined as being symmetric when

$$\mathbf{A}' = \mathbf{A}\,.$$

That can occur only when **A** is square. Its rows are then mirror images of its columns:

$$\begin{bmatrix} 1 & 7 & 0 \\ 7 & 2 & -3 \\ 0 & -3 & 9 \end{bmatrix}' = \begin{bmatrix} 1 & 7 & 0 \\ 7 & 2 & -3 \\ 0 & -3 & 9 \end{bmatrix};$$

and $a_{ij} = a_{ji}$.

BB′ and **B**′**B** are both symmetric. This is true for any **B**. Then **BB**′ (and **B**′**B**) have diagonal elements that are sums of squares of elements of rows (columns) of **B**: and

$$\text{tr}(\mathbf{BB}') = \text{tr}(\mathbf{B}'\mathbf{B}) = \Sigma_i \Sigma_j b_{ij}^2\,.$$

When **B** is real, $\mathbf{BB}' = \mathbf{0}$ and $\text{tr}(\mathbf{BB}') = 0$ each imply $\mathbf{B} = \mathbf{0}$.

Elementary vectors

Columns of identity matrices are *elementary vectors*, represented as $\mathbf{e}_i^{(n)}$, the i'th column in **I** of order n.

Skew-symmetric matrices

$\mathbf{A}' = -\mathbf{A}$; defines **A** as *skew-symmetric.*

Summing vectors

A vector having every element a one (1.0) is a *summing vector*, often denoted as **1**. It is so named because $\mathbf{1}'\mathbf{x}$ is the sum of all elements in **x**.

Matrices having every element unity

$\mathbf{J}_{p\times k} = \mathbf{1}_p\mathbf{1}'_k$ is a matrix having every element being 1.0. Its most frequent occurrence in statistics is when it is square, $\mathbf{J}_n = \mathbf{1}_n\mathbf{1}'_n$. A useful variant is $\bar{\mathbf{J}}_n = (1/n)\mathbf{J}_n$. Then

$$\mathbf{C}_n = \mathbf{I}_n - \bar{\mathbf{J}}_n \, .$$

is a *centering matrix* with

$$\mathbf{C}_n\mathbf{x} = \{x_i - \bar{x}\} \qquad \text{and} \qquad \mathbf{x}'\mathbf{C}_n\mathbf{x} = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

for $\bar{x} = \sum_{i=1}^{n} x_i/n = \mathbf{1}'\mathbf{x}/n$.

Probability transition matrices

When elements of a matrix $\mathbf{P}$ are probabilities that add to unity over each row $\mathbf{P1} = \mathbf{1}$. Then $\mathbf{P}^k\mathbf{1} = \mathbf{1}$ for any positive integer k, and $\mathbf{P}$ is called a *probability transition matrix.* It is *doubly stochastic* if $\mathbf{1}'\mathbf{P} = \mathbf{1}'$ also.

Idempotent matrices

$\mathbf{A}$ is *idempotent* when $\mathbf{A}^2 = \mathbf{A}$; then $\mathbf{I} - \mathbf{A}$ is idempotent also (but $\mathbf{A} - \mathbf{I}$ is not).

Orthogonality

The *norm* of a real vector $\mathbf{x}$ is $\sqrt{\mathbf{x}'\mathbf{x}}$.

$\mathbf{x}$ is a *unit vector* when $\mathbf{x}'\mathbf{x} = 1$.

$\mathbf{u} = (\mathbf{x}'\mathbf{x})^{-\frac{1}{2}}\mathbf{x}$ is always a unit vector.

Non-null vectors $\mathbf{x}$ and $\mathbf{y}$ are *orthogonal vectors* when $\mathbf{x}'\mathbf{y} = 0$ $(= \mathbf{y}'\mathbf{x})$.

Vectors $\mathbf{v}$ and $\mathbf{w}$ are *orthonormal vectors* when they are orthogonal $(\mathbf{v}'\mathbf{w} = 0)$ and each is a unit vector $(\mathbf{v}'\mathbf{v} = 1$ and $\mathbf{w}'\mathbf{w} = 1)$.

A collection of vectors of the same order is said to be an *orthogonal set of vectors* when they are pairwise orthonormal.

When $\mathbf{P}_{r\times c}$ has rows that are an orthonormal set, $\mathbf{PP}' = \mathbf{I}$. If $\mathbf{P}$ is square with orthonormal rows (columns) then its columns (rows) are orthonormal also, $\mathbf{PP}' = \mathbf{I} = \mathbf{P}'\mathbf{P}$ and $\mathbf{P}$ is an *orthogonal matrix.*

Certain special forms of orthogonal matrices go by the names Helmert, Givens and Householder. The latter, for example, is $\mathbf{I} - 2\mathbf{hh}'$ when $\mathbf{h}'\mathbf{h} = 1$.

Quadratic forms

$\mathbf{x}'\mathbf{A}\mathbf{x}$ is a *quadratic form*, in which $\mathbf{A}$ can always be (taken as) symmetric. $\mathbf{x}'\mathbf{A}\mathbf{x}$ is a homogeneous second order function of the elements of $\mathbf{x}$:

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_i x_i^2 a_{ii} + \sum x_i x_j a_{ij} = \sum_i x_i^2 a_{ii} + \sum_{j \neq i} x_i x_j (a_{ij} + a_{ji})$$

and on taking $\mathbf{A} = \mathbf{A}'$, i.e., $a_{ij} = a_{ji}$,

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_i x_i^2 a_{ii} + 2\sum_{j>i} x_i x_j a_{ij} \,.$$

If $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x}'\mathbf{A}\mathbf{x}$ is called a *positive definite (p.d.) quadratic form*, and $\mathbf{A}$ $(=\mathbf{A}')$ is a *p.d. matrix*. If $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \neq 0$ *and* $\mathbf{x}'\mathbf{A}\mathbf{x} = 0$ for some $\mathbf{x} \neq 0$, then $\mathbf{x}'\mathbf{A}\mathbf{x}$ and $\mathbf{A}$ are *positive semi-definite* (p.s.d.). The classes of quadratic forms and matrices that includes those which are p.d. and p.s.d. are called *non-negative definite* (n.n.d.).

DETERMINANTS

Definition

Associated with any square matrix $\mathbf{A}_{n \times n}$ is its determinant $|\mathbf{A}|$. It is a scalar, an n-order, homogeneous polynomial function of the elements. Two easy examples are for $\mathbf{A}$ of order 2 and 3:

$$|\mathbf{X}| = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = a_1 b_2 - a_2 b_1$$

and

$$|\mathbf{Y}| = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = a_1 b_2 c_3 + a_2 b_3 c_1 + a_3 b_1 c_3 - a_3 b_2 c_1 - a_1 b_3 c_2 - a_2 b_1 c_3 \,.$$

For $\mathbf{A}$ of order n the definition is more difficult: $|\mathbf{A}|$ is the sum of the n different terms that are each a signed product of one element from every row and column of $\mathbf{A}$. In writing $|\mathbf{A}|$ with rows being $\mathbf{a}'$, $\mathbf{b}'$, $\mathbf{c}'$, $\cdots$, a product written in alphabetic order has sign equal to $(-1)^p$ with p being the sum of the number of reverse sequences of the subscripts. For example, $a_2 b_3 c_1$ in the preceding $|\mathbf{Y}|$ has $p = 2$ because 2,1 and 3,1 are reverse sequences; hence the sign for $a_2 b_3 c_1$ is $(-1)^2 = +1$. And for $a_3 b_2 c_1$ there are three reverse sequences, 3,2 and 3,1 and 2,1 and so the sign is $(-1)^3 = -1$.

Minors and Cofactors

Deleting from $|\mathbf{A}|$ the row and column containing a_{ij} leaves a determinant of order $n-1$ that is called the *minor*, $|\mathbf{M}_{ij}|$, of a_{ij} in $|\mathbf{A}|$. And $(-1)^{i+j}|\mathbf{M}_{ij}|$, the *signed minor*, is called the *cofactor* of a_{ij} in $|\mathbf{A}|$: $c_{ij} = (-1)^{i+j}|\mathbf{M}_{ij}|$. Then

$$|\mathbf{A}| = \sum_{i=1}^{n} a_{ij}c_{ij} \quad \forall\ j = \sum_{j=1}^{n} a_{ij}c_{ij} \quad \forall\ i$$

but

$$0 = \sum_{i=1}^{n} a_{ij}c_{ij'} \quad \forall\ j \neq j' = \sum_{j=1}^{n} a_{ij}c_{i'j} \quad \forall\ i \neq i' .$$

Calculation

Computers now handle the calculation of arithmetic determinants. Numerous available shortcuts and associated properties of determinants are detailed in the literature, which was especially rich on this subject up through the 1930s. Searle (1987) deals with a few of these topics.

Some properties useful for statistics

$$|\mathbf{A}'| = |\mathbf{A}| .$$

$$|\mathbf{A}^k| = (|\mathbf{A}|)^k \qquad \text{for integer } k .$$

$$|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}| .$$

$$|\mathbf{A}| = +1 \qquad \text{for orthogonal } \mathbf{A} .$$

$$|\mathbf{A}| = 0 \qquad \text{for idempotent } \mathbf{A}\text{, except for } \mathbf{A} = \mathbf{I} .$$

$$|\mathbf{I}| = 1 .$$

$$|\lambda\mathbf{A}_{n \times n}| = \lambda^n\ |\mathbf{A}| \qquad \text{for scalar } \lambda .$$

INVERSE MATRICES

Existence

In matrix arithmetic the very definition of multiplication precludes any obvious definition of division. Indeed, there is no such thing as matrix division; *division by a matrix does not exist.* Instead, multiplication by an inverse matrix is used, similar to the scalar equivalence of dividing by six (for example) being identical to multiplying by $1/6 = 6^{-1}$, the inverse of six. And

$$(6^{-1})6 = 1 = 6(6^{-1}) .$$

There is one big difference: whereas every scalar has an inverse, not every matrix does.

Suppose $\mathbf{A}$ has an inverse. Denote it by $\mathbf{A}^{-1}$, as is customary. Then with $\mathbf{I}$ being a "one" of matrix algebra, the matrix analogy of scalars is $(\mathbf{A}^{-1})\mathbf{A} = \mathbf{I} = \mathbf{A}(\mathbf{A}^{-1})$ where the parentheses are solely for emphasis, the usual writing being

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

This requirement demands that two conditions must be satisfied in order for $\mathbf{A}^{-1}$ to exist:

(a) $\mathbf{A}$ must be square .

(b) $|\mathbf{A}| \neq 0$.

If either or both (a) and (b) are not satisfied $\mathbf{A}$ has no inverse; note, particularly, that every rectangular matrix has no inverse.

When $|\mathbf{A}| \neq 0$, $\mathbf{A}$ is called *non-singular*, and if $|\mathbf{A}| = 0$ $\mathbf{A}$ is called *singular.*

Form

The general form of $\mathbf{A}^{-1}$ is

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\begin{bmatrix}\text{The matrix which is } \mathbf{A} \\ \text{with every element} \\ \text{replaced by its cofactor}\end{bmatrix}^{\text{transposed}} .$$

and $|\mathbf{A}|\mathbf{A}^{-1}$ is called the *adjugate* or *adjoint* of $\mathbf{A}$.

Some basic properties

(i) $\mathbf{A}^{-1}$ is unique (for given $\mathbf{A}$) .

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

(iii) $\mathbf{A}^{-1}$ is non-singular.

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

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

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

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

In all of these results, and whenever an inverse is used, one must always be certain that the matrix satisfies (a) and (b) above, namely squareness and non-zero determinant.

Four special cases

Denote a diagonal matrix having all its diagonal elements $\lambda_1, \cdots, \lambda_n$ non-zero by

$$\mathbf{D} = \left\{_d\ \lambda_i\right\}_{i=1}^{n}; \quad \text{then} \quad \mathbf{D}^{-1} = \left\{_d\ 1/\lambda_i\right\}_{i=1}^{n}.$$

$$\mathbf{I}^{-1} = \mathbf{I}.$$

$$(\mathrm{a}\mathbf{I}_n + \mathrm{b}\mathbf{J}_n)^{-1} = \tfrac{1}{\mathrm{a}}\left(\mathbf{I}_n - \frac{\mathrm{b}}{\mathrm{a}+n\mathrm{b}}\mathbf{J}_n\right).$$

$$\mathbf{PP}' = \mathbf{I} = \mathbf{P}'\mathbf{P} \quad \text{implies} \quad \mathbf{P}^{-1} = \mathbf{P}'.$$

Algebra with inverses

Compared to using division in scalar algebra one has to be much more careful in using inverses in matrix algebra. This is because one never divides by a matrix; instead, in dealing with equations, one multiplies by an inverse. For example, given $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{AX} = \mathbf{B}$, the equation can be pre-multiplied, on both sides, by $\mathbf{A}^{-1}$ (providing it exists) to get $\mathbf{A}^{-1}\mathbf{AX} = \mathbf{A}^{-1}\mathbf{B}$ and thus $\mathbf{IX} = \mathbf{A}^{-1}\mathbf{B}$ or $\mathbf{X} = \mathbf{A}^{-1}\mathbf{B}$. And note that $\mathbf{X}$ does not equal $\mathbf{BA}^{-1}$. Providing conformability is satisfied one could post-multiply $\mathbf{AX} = \mathbf{B}$ by $\mathbf{A}^{-1}$ and get $\mathbf{AXA}^{-1} = \mathbf{BA}^{-1}$; but that is it. No further simplification occurs.

Suppose we have $\mathbf{P}$, $\mathbf{Q}$ and $\mathbf{K}$ such that $\mathbf{PK} = \mathbf{QK}$. This leads to $\mathbf{P} = \mathbf{Q}$ *only* if $\mathbf{K}^{-1}$ exists.

Inverses can also be used in factoring; for example, $\mathbf{R} + \mathbf{RST} = \mathbf{R}(\mathbf{I} + \mathbf{ST}) = \mathbf{R}(\mathbf{T}^{-1} + \mathbf{S})\mathbf{T}$, provided $\mathbf{T}^{-1}$ exists.

Verifying the form of a particular inverse is often achieved by the following argument. Suppose it is postulated that $\mathbf{A}$ inverse is $\mathbf{Q}$. Verifying this can be achieved by considering the product $\mathbf{AQ}$. If that can be shown equal to $\mathbf{I}$, thus $\mathbf{AQ} = \mathbf{I}$, then $\mathbf{A}^{-1}\mathbf{AQ} = \mathbf{A}^{-1}\mathbf{I}$, i.e., $\mathbf{Q} = \mathbf{A}^{-1}$. For example, suppose $\mathbf{A}$ is $(\mathbf{I} + \mathbf{XY})$ and $\mathbf{Q}$ is $\mathbf{I} - \mathbf{X}(\mathbf{I} + \mathbf{YX})^{-1}\mathbf{Y}$. $\mathbf{A}^{-1}$ is shown to be $\mathbf{Q}$ by considering $\mathbf{AQ}$:

$$\begin{aligned}
\mathbf{AQ} &= (\mathbf{I} + \mathbf{XY})[\mathbf{I} - \mathbf{X}(\mathbf{I} + \mathbf{YX})^{-1}\mathbf{Y}] \\
&= \mathbf{I} + \mathbf{XY} - (\mathbf{I} + \mathbf{XY})\mathbf{X}(\mathbf{I} + \mathbf{YX})^{-1}\mathbf{Y} \\
&= \mathbf{I} + \mathbf{XY} - (\mathbf{X} + \mathbf{XYX})(\mathbf{I} + \mathbf{YX})^{-1}\mathbf{Y} \\
&= \mathbf{I} + \mathbf{XY} - \mathbf{X}(\mathbf{I} + \mathbf{YX})(\mathbf{I} + \mathbf{YX})^{-1}\mathbf{Y} \\
&= \mathbf{I} + \mathbf{XY} - \mathbf{XY} \\
&= \mathbf{I},
\end{aligned}$$

and so $\mathbf{A}^{-1} = \mathbf{Q}$.

Computers and inverses

The arithmetic required for calculating an inverse matrix can be voluminous. Fortunately, computers have eased this situation enormously and many software packages include reliable routines for doing the arithmetic. Nevertheless, there are cases where rounding error can lead to erroneous results; thankfully, this occurs very very seldom, and software often handles it satisfactorily.

RANK

Linear dependence and independence of vectors

$$\mathbf{Xa} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_c] \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_i \\ \vdots \\ a_c \end{bmatrix} = \sum_{i=1}^{c} a_i \mathbf{x}_i$$

is a vector. It is a *linear combination* of the vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_c$.

Given $\mathbf{X}$ (with all columns non-null), if a non-null vector $\mathbf{a}$ exists such that $\mathbf{Xa} = \mathbf{0}$, then the columns of $\mathbf{A}$ are said to be a set of *linearly dependent vectors.* If no such $\mathbf{a}$ exists, the columns are *linearly independent vectors.* These definitions exclude null vectors.

A definition of rank

If c columns are linearly dependent there is always a smaller number of them that are linearly independent. In fact, there may be several sets of less than c columns that are linearly independent, with those sets not necessarily all having the same number of columns. The greatest number of columns in such a set is called the *rank of* $\mathbf{A}$, often denoted $r(\mathbf{A})$. Thus $r(\mathbf{A})$ is the largest number of linearly independent columns available from $\mathbf{A}$. The "largest" is usually omitted. Thus $r(\mathbf{A})$ is the number of linearly independent columns in $\mathbf{A}$.

Some properties and consequences

Rank is an important and exceedingly useful concept in matrix algebra, with widespread applications. A list of some of the properties of rank follows.

(a) The numbers of linearly independent rows and columns in a matrix are the same, $r(\mathbf{A})$.

(b) $r(\mathbf{0}) = 0$.

(c) $r(\mathbf{A}_{p \times q}) \leq p$ and $r(\mathbf{A}_{p \times q}) \leq q$.

(d) $r(\mathbf{A}_{n \times n}) \leq n$.

(e) $r(\mathbf{A}_{n \times n}) < n \Leftrightarrow$ **A** is singular, $|\mathbf{A}| = \mathbf{0}$, with $\mathbf{A}^{-1}$ not existing.

(f) $r(\mathbf{A}_{n \times n}) = n \Leftrightarrow$ **A** is non-singular, $|\mathbf{A}| \neq \mathbf{0}$, with $\mathbf{A}^{-1}$ existing: **A** is said to be of *full rank.*

(g) $r(\mathbf{A}_{p \times q}) = p < q$ means **A** has *full row rank.*

(h) $r(\mathbf{A}_{p \times q}) = q < p$ means **A** has *full column rank.*

(i) $\mathbf{A}_{p \times q}$ having rank r can always be expressed as $\mathbf{A}_{p \times q} = \mathbf{K}_{p \times r}\mathbf{L}_{r \times q}$ where **K** has full column rank r and **L** has full row rank r.

(j) $r(\mathbf{AB}) \leq$ lesser of $r(\mathbf{A})$ and $r(\mathbf{B})$.

(k) $r(\mathbf{A}) = tr(\mathbf{A})$ for idempotent **A**.

(l) $r(\mathbf{A}) = r(\mathbf{A}')$.

(m) $r(\mathbf{A}) = r(\mathbf{AA}')$.

(n) $r(\mathbf{A}) = r(\mathbf{TA})$ for non-singular **T**.

(o) $r(\mathbf{A}^{-1}) = r(\mathbf{A}) = n$ for $\mathbf{A}_{n \times n}$

Left and right inverses

For given $\mathbf{A}_{r \times c}$ there exists

(i) $\mathbf{A}^{-1}$, the inverse of **A**, such that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I} = \mathbf{AA}^{-1}$ if and only if **A** is square, with $|\mathbf{A}| \neq 0$;

or (ii) $\mathbf{L}_{c \times r}$, a *left inverse* of **A**, such that $\mathbf{LA} = \mathbf{I}_c$ (and $\mathbf{AL} \neq \mathbf{I}_r$) only if **A** has full column rank;

or (iii) $\mathbf{R}_{c \times r}$, a *right inverse* of **A**, such that $\mathbf{AR} = \mathbf{I}_r$ (and $\mathbf{RA} \neq \mathbf{I}_c$) only if **A** has full row rank;

or (iv) Neither an $\mathbf{A}^{-1}$, **L** *nor* **R** of (i), (ii) or (iii); e.g., any matrix having at least one null row and column.

Only when $\mathbf{A}^{-1}$ exists does **A** have both an **L** and an **R**; and they both equal $\mathbf{A}^{-1}$. Otherwise, if **A** has full column (row) rank it has left (right) inverses of many values.

Vector spaces

Since a vector of order n has n elements, it can be considered as a point in n-space, which is denoted R^n. Consider a set of vectors S, in R^n. Suppose, for every pair of vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ in S, that both the sum $\mathbf{x}_i + \mathbf{x}_j$ and the vectors $a\mathbf{x}_i$ and $b\mathbf{x}_i$ for any scalars a and b are in S; then S is a *vector space.*

Suppose every vector in the vector space S can be expressed as a linear combination of the set of t vectors, $\mathbf{x}_i, \mathbf{x}_2, \cdots, \mathbf{x}_i$. Then that set *spans*, or *generates*, S and is called a *spanning* set of S. If those t vectors are also linearly independent they are said to be *a basis* for S and the number of such vectors is the *dimension* of S, dim(S).

There are many vector spaces of order n; and each of them usually has several bases.

Range and null spaces

A of rank r has r linearly independent columns. All vectors that are linear combinations of those columns form a vector space. It is known as the *column space* of **A**, the *range* of **A** or the *manifold* of **A**, often denoted by $\mathcal{R}(\mathbf{A})$. Clearly $r = r(\mathbf{A}) = \dim[\mathcal{R}(\mathbf{A})]$.

The space defined by the many vectors **x** for which $\mathbf{Ax} = \mathbf{0}$ (with **A** being rectangular or square and singular) is the *null space* of **A**, denoted $\mathcal{N}(\mathbf{A})$. Its dimension is the *nullity* of **A**: nullity $(\mathbf{A}) = \dim[\mathcal{N}(\mathbf{A})]$.

EQUIVALENT and CONGRUENT CANONICAL FORMS

Elementary operators

Three particular adaptations of identity matrices are *elementary operators; each is an identity matrix* with (i) two rows (or columns) interchanged, or (ii) λ in place of a one in the diagonal, or (iii) λ in place of a zero in an off-diagonal element. These and all products of any numbers of them are non-singular.

Equivalent canonical form

For any $\mathbf{A}_{p \times q}$, of rank r, there always exists a **P** and a **Q**, each a product of elementary operators, such that

$$\mathbf{P}_{p\times p}\mathbf{A}_{p\times q}\mathbf{Q}_{q\times q} = \begin{bmatrix} \mathbf{I}_{r\times r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} = \mathbf{K}, \text{ say} .$$

K is the *equivalent canonical form* of **A**; or the *canonical form under equivalence* of **A**. Because **P** and **Q** are products of elementary operators, they are non-singular, and so the equation $\mathbf{PAQ} = \mathbf{K}$ leads to $\mathbf{A} = \mathbf{P}^{-1}\mathbf{K}\mathbf{Q}^{-1}$. If **A** is non-singular, $\mathbf{K} = \mathbf{I}$ and $\mathbf{A}^{-1} = \mathbf{QP}$.

Congruent canonical form

When **A** is symmetric (and hence square), the **Q** of **PAQ** can be **P′** giving

$$\mathbf{PAP}' = \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} = \mathbf{C}$$

known as the *congruent canonical form* of **A** or the *canonical form under congruence*.

En route to deriving **C** one can obtain the form

$$\mathbf{P}_*\mathbf{AP}'_* = \begin{pmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where $\mathbf{D}_r$ is a diagonal matrix of order and rank r. For **A** being real, $\mathbf{P}_*$ will be real; but if $\mathbf{D}_r$ has negative elements, **P** in obtaining **C** will be complex. For **A** being non-negative definite, elements of $\mathbf{D}_r$ are always positive and **P** is always real.

Utility: sums of squares

The utility of these canonical forms is their existence. For each **A** there are many values of **P** (and **Q**) but usually not any one of them is of particular interest. It is the fact that they exist that is important, and which provides the means for establishing other useful results. For example, consider the quadratic form $q = \mathbf{x}'\mathbf{Ax}$ with $\mathbf{A} = \mathbf{A}'$ of rank r. Then there is a **P** such that $\mathbf{PAP}' = \mathbf{C}$. Thus $q = \mathbf{x}'\mathbf{P}^{-1}\mathbf{PAP}'(\mathbf{P}')^{-1}\mathbf{x}$ and letting $\mathbf{y} = (\mathbf{P}')^{-1}\mathbf{x}$ gives $q = \mathbf{y}'\mathbf{Cy}$ which, because $\mathbf{C} = \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$ becomes $q = \sum_{i=1}^{r} y_i^2$. Thus, without knowing **P** except for its existence and non-singularity, we can show that a quadratic form can always be expressed as a sum of r squared terms where r is the rank of the (symmetric) matrix **A** of the quadratic form. That is a result of great importance in considering the distribution of quadratic forms of normally distributed random variables.

GENERALIZED INVERSES

Definition

For any non-null matrix $\mathbf{A}$, there is a unique matrix $\mathbf{M}$ satisfying

$$\text{(i)}\quad \mathbf{AMA} = \mathbf{A}\,, \qquad\qquad \text{(ii)}\quad \mathbf{MAM} = \mathbf{M}\,,$$

$$\text{(iii)}\quad (\mathbf{AM})' = \mathbf{AM}\,, \qquad \text{and} \qquad \text{(iv)}\quad (\mathbf{MA})' = \mathbf{MA}\,.$$

These are the *Penrose* conditions and $\mathbf{M}$ is the *Moore-Penrose inverse.* Whereas $\mathbf{M}$ is unique, there are (with one exception) many matrices $\mathbf{G}$ satisfying

$$\mathbf{AGA} = \mathbf{A}\,,$$

which is condition (i). Each matrix $\mathbf{G}$, satisfying $\mathbf{AGA} = \mathbf{A}$ is called a *generalized inverse* of $\mathbf{A}$, and if it also satisfies $\mathbf{GAG} = \mathbf{G}$ it is a *reflexive generalized inverse.* The exception is when $\mathbf{A}$ is non-singular: there is then only one $\mathbf{G}$, namely $\mathbf{G} = \mathbf{A}^{-1}$.

Arbitrariness

That there are many matrices $\mathbf{G}$ can be illustrated by showing ways in which from one $\mathbf{G}$ others can be obtained. Thus if $\mathbf{A}$ is partitioned as

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

where $\mathbf{A}_{11}$ is non-singular with the same rank as $\mathbf{A}$, then

$$\mathbf{G} = \begin{bmatrix} \mathbf{A}_{11}^{-1} - \mathbf{U}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} - \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{V} - \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{W}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{bmatrix}$$

is a generalized inverse of $\mathbf{A}$ for any values of $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$. This can be used to show that a generalized inverse of a symmetric matrix is not necessarily symmetric; and that of a singular matrix is not necessarily singular (Searle, 1982, p. 219).

A simpler illustration of arbitrariness is that if $\mathbf{G}$ is a generalized inverse of $\mathbf{A}$ then so is

$$\mathbf{G} = \mathbf{GAG} + (\mathbf{I} - \mathbf{GA})\mathbf{S} + \mathbf{T}(\mathbf{I} - \mathbf{AG})$$

for any values of $\mathbf{S}$ and $\mathbf{T}$.

Generalized inverses of $\mathbf{X'X}$

The matrix $\mathbf{X'X}$ plays an important role in statistics, usually involving a generalized inverse thereof, which has several useful properties. Thus for $\mathbf{G}$ satisfying

$$\mathbf{X'XGX'X} = \mathbf{X'X} ,$$

$\mathbf{G}'$ is also a generalized inverse of $\mathbf{X'X}$ (and $\mathbf{G}$ is not necessarily symmetric). Also

$\mathbf{XGX'X} = \mathbf{X}$.

$\mathbf{XGX}'$ is invariant to $\mathbf{G}$.

$\mathbf{XGX}'$ is symmetric, whether $\mathbf{G}$ is or not .

$\mathbf{XGX}' = \mathbf{XX}^{+}$ for $\mathbf{X}^{+}$ being the Moore-Penrose inverse of $\mathbf{X}$.

SOLVING LINEAR EQUATIONS

A single solution

Given $\mathbf{A}$ and $\mathbf{y}$, equations $\mathbf{Ax} = \mathbf{y}$ are linear in the unknowns, the elements of $\mathbf{x}$. When $\mathbf{A}$ is non-singular the equations are solved uniquely as $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$. But for singular or rectangular $\mathbf{A}$ solutions involve using a generalized inverse of $\mathbf{A}$. The following results apply.

First, equations $\mathbf{Ax} = \mathbf{y}$ are said to be *consistent* when any linear relationships existing among rows of $\mathbf{A}$ also exist among elements of $\mathbf{y}$. Only then do solutions exist. And for singular or rectangular $\mathbf{A}$ there will be many solutions for $\mathbf{x}$, except when $\mathbf{A}$ has full column rank, whereupon there is only one solution, $\mathbf{x} = (\mathbf{A'A})^{-1}\mathbf{A'y}$. And this includes, of course, the case of non-singular $\mathbf{A}$.

Many solutions

When $\mathbf{A}$ has less than full column rank, there are many solutions. They are characterized as follows, with $\mathbf{G}$ being a generalized inverse satisfying $\mathbf{AGA} = \mathbf{A}$.

1. $\tilde{\mathbf{x}} = \mathbf{Gy}$ is a solution if and only if $\mathbf{AGA} = \mathbf{A}$.

2. $\tilde{\mathbf{x}} = \mathbf{Gy} + (\mathbf{I} - \mathbf{GA})\mathbf{z}$ is a solution for any arbitrary $\mathbf{z}$ of the same order as $\mathbf{x}$.

3. Letting $\mathbf{G}$ take all its possible values in $\tilde{\mathbf{x}} = \mathbf{Gy}$ (for $\mathbf{y} \neq 0$) generates all possible solutions.

4. For a given $\mathbf{G}$, letting $\mathbf{z}$ take all possible values in $\tilde{\mathbf{x}} = \mathbf{Gy} + (\mathbf{I} - \mathbf{GA})\mathbf{z}$ generates all possible solutions.

5. When $\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \cdots, \tilde{\mathbf{x}}_t$ are any solutions, $\sum_{i=1}^{t} \lambda_i \tilde{\mathbf{x}}_i$ is a solution (with $\mathbf{y} \neq \mathbf{0}$) if and only if $\sum_{i=1}^{t} \lambda_i = 1$; this condition is not needed when $\mathbf{y} = \mathbf{0}$.

6. For $\mathbf{A}_{p \times q}$ and $\mathbf{y} \neq \mathbf{0}$ there are $q - r(\mathbf{A}) + 1$ $[q - r(\mathbf{A})$ when $\mathbf{y} = \mathbf{0}]$ linearly independent solutions.

7. The value of $\mathbf{k}'\tilde{\mathbf{x}}$ is invariant to $\tilde{\mathbf{x}}$ if and only if $\mathbf{k}' = \mathbf{k}'\mathbf{G}\mathbf{A}$.

8. When $\mathbf{y} = \mathbf{0}$, solutions are orthogonal to rows of $\mathbf{A}$; and solutions orthogonal to each other can always be derived. The vector space spanned by the solutions, sometimes called the *solution space*, is the *orthogonal complement* of the row space of $\mathbf{A}$.

PARTITIONED MATRICES

Some results for partitioned matrices that get used in statistics are as follows.

Orthogonality

If $\mathbf{P} = [\mathbf{A} \quad \mathbf{B}]$ is orthogonal

$$\mathbf{P}\mathbf{P}' = \mathbf{I} \;\Rightarrow\; [\mathbf{A} \quad \mathbf{B}]\begin{bmatrix} \mathbf{A}' \\ \mathbf{B}' \end{bmatrix} = \mathbf{I} \;\Rightarrow\; \mathbf{A}\mathbf{A}' + \mathbf{B}\mathbf{B}' = \mathbf{I}$$

$$\mathbf{P}'\mathbf{P} = \mathbf{I} \;\Rightarrow\; \begin{bmatrix} \mathbf{A}' \\ \mathbf{B}' \end{bmatrix}[\mathbf{A} \quad \mathbf{B}] = \mathbf{I} \;\Rightarrow\; \begin{bmatrix} \mathbf{A}'\mathbf{A} & \mathbf{A}'\mathbf{B} \\ \mathbf{B}'\mathbf{A} & \mathbf{B}'\mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

$$\Rightarrow\; \mathbf{A}'\mathbf{A} = \mathbf{I}, \quad \mathbf{A}'\mathbf{B} = \mathbf{0} \quad \text{and} \quad \mathbf{B}'\mathbf{B} = \mathbf{I}.$$

Note: $\mathbf{A}\mathbf{A}'$ and $\mathbf{B}\mathbf{B}'$ are not identity matrices .

Determinants

$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = |\mathbf{A}|\,|\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}| = |\mathbf{D}|\,|\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}|,$$

providing $\mathbf{A}^{-1}$ and $\mathbf{D}^{-1}$ exist, where needed.

Inverses

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

$$= \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}^{-1} \end{bmatrix} + \begin{bmatrix} \mathbf{I} \\ -\mathbf{D}^{-1}\mathbf{C} \end{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} [\mathbf{I} \quad -\mathbf{B}\mathbf{D}^{-1}] ,$$

again providing $\mathbf{A}^{-1}$ and $\mathbf{D}^{-1}$ exist as needed.

Schur complements

In $\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$ the *Schur complement* of **A** is $\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ and that of **D** is $\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$. The inverse of one involves that of the other:

$$(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} = \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} .$$

This results also applies when the two minus signs are changed to plus, and the plus to minus. It also has some useful special cases, e.g.,

$$(\mathbf{D} \pm \lambda\mathbf{t}\mathbf{t}')^{-1} = \mathbf{D}^{-1} \mp \frac{\mathbf{D}^{-1}\mathbf{t}\mathbf{t}'\mathbf{D}}{(\lambda^{-1} \pm \mathbf{t}'\mathbf{D}^{-1}\mathbf{t})} .$$

Generalized inverses

By analogy with expressions for the inverse one might expect

$$\mathbf{Q}^* = \begin{bmatrix} \mathbf{A}^- & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\mathbf{A}^-\mathbf{B} \\ \mathbf{I} \end{bmatrix} (\mathbf{D} - \mathbf{C}\mathbf{A}^-\mathbf{B})^- [-\mathbf{C}\mathbf{A}^- \quad \mathbf{I}]$$

to be a generalized inverse of

$$\mathbf{Q} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} .$$

It is, if and only if $r(\mathbf{Q}) = r(\mathbf{A}) + r(\mathbf{D} - \mathbf{C}\mathbf{A}^-\mathbf{B})$. Satisfying this rank condition depends upon $\mathbf{A}^-$. For some values of $\mathbf{A}^-$ the condition will be satisfied and for others it will not. Only when it is satisfied will $\mathbf{Q}^*$ be a generalized inverse of **Q**.

Direct sums

The direct sum of matrices **A** and **B**, each of any order is defined as

$$\mathbf{A} \oplus \mathbf{B} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} .$$

Extension to the the direct sum of more than two matrices is straightforward.

Provided the needed conformability requirements are met

$$(\mathbf{A} \oplus \mathbf{B}) + (\mathbf{C} \oplus \mathbf{D}) = (\mathbf{A} + \mathbf{C}) \oplus (\mathbf{B} + \mathbf{D}) \, ,$$

$$(\mathbf{P} \oplus \mathbf{Q})(\mathbf{L} \oplus \mathbf{M}) = \mathbf{PL} \oplus \mathbf{QM} \, ,$$

and

$$(\mathbf{X} \oplus \mathbf{Y})^{-1} = \mathbf{X}^{-1} \oplus \mathbf{Y}^{-1} \, .$$

Direct products

The *direct product* of two matrices each of any order is defined as

$$\mathbf{A}_{p \times q} \otimes \mathbf{B}_{m \times n} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1q}\mathbf{B} \\ \vdots & & \\ a_{p1}\mathbf{B} & \cdots & a_{pq}\mathbf{B} \end{bmatrix}_{pm \times qn} = \{a_{ij}\mathbf{B}\}_{i=1,\, j=1}^{p \quad q} \, .$$

It is sometimes called the *Kronecker product.* Some properties follow — assuming conformability requirements are met.

1. $\mathbf{x}' \otimes \mathbf{y} = \mathbf{y}\mathbf{x}' = \mathbf{y} \otimes \mathbf{x}'$.
2. $\lambda \otimes \mathbf{A} = \lambda\mathbf{A} = \mathbf{A} \otimes \lambda$.
3. $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$; *not* $\mathbf{B}' \otimes \mathbf{A}'$.
4. $(\mathbf{A} \otimes \mathbf{B})(\mathbf{X} \otimes \mathbf{Y}) = \mathbf{AX} \otimes \mathbf{BY}$.
5. $(\mathbf{P} \otimes \mathbf{Q})^{-1} = \mathbf{P}^{-1} \otimes \mathbf{Q}^{-1}$; *not* $\mathbf{Q}^{-1} \otimes \mathbf{P}^{-1}$.
6. $[\mathbf{A}_1 \quad \mathbf{A}_2] \otimes \mathbf{B} = [\mathbf{A}_1 \otimes \mathbf{B} \quad \mathbf{A}_2 \otimes \mathbf{B}]$

   $\mathbf{A} \otimes [\mathbf{B}_1 \quad \mathbf{B}_2] \neq [\mathbf{A} \otimes \mathbf{B}_1 \quad \mathbf{A} \otimes \mathbf{B}_2]$.
7. $r(\mathbf{A} \otimes \mathbf{B}) = r(\mathbf{A})r(\mathbf{B})$.
8. $tr(\mathbf{A} \otimes \mathbf{B}) = tr(\mathbf{A})tr(\mathbf{B})$.
9. $|\mathbf{A}_{p \times p} \otimes \mathbf{B}_{m \times m}| = |\mathbf{A}|^m |\mathbf{B}|^p$.

Sometimes $\mathbf{A} \otimes \mathbf{B} = \{a_{ij}\mathbf{B}\}$ as defined above is called the *right direct product* to distinguish it from $\mathbf{B} \otimes \mathbf{A}$, which is then called the *left direct product*; and on rare occasion $\{a_{ij}\mathbf{B}\}$ will be found defined as $\mathbf{B} \otimes \mathbf{A}$.

EIGEN ROOTS AND VECTORS

The equation

$$\mathbf{Au} = \lambda\mathbf{u}, \qquad \text{i.e.,} \qquad (\mathbf{A} - \lambda\mathbf{I})\mathbf{u} = \mathbf{0}$$

has solutions for $\mathbf{u}$ provided $\mathbf{A} - \lambda\mathbf{I}$ is singular. This occurs when

$$|\mathbf{A} - \lambda\mathbf{I}| = \mathbf{0}\,.$$

This is called the *characteristic equation* of $\mathbf{A}$; for $\mathbf{A}_{n \times n}$ it is a polynomial of order n and therefore has n solutions for λ. Those solutions are the *eigenroots* of $\mathbf{A}$. They can be real or complex, positive or negative, or zero. For each eigenroot, λ_* say, a corresponding value of $\mathbf{u}$ can be obtained from solving the equations $(\mathbf{A} - \lambda_*\mathbf{I})\mathbf{u} = \mathbf{0}$ as

$$\mathbf{u}_* = [\mathbf{I} - (\mathbf{A} - \lambda_*\mathbf{I})^{-}(\mathbf{A} - \lambda_*\mathbf{I})]\mathbf{z}$$

for arbitrary $\mathbf{z}$. (Searle, 1982, Section 11.4 has details.) $\mathbf{u}_*$ is the *eigenvector* corresponding to λ_*.

Numerical example

For $\mathbf{A} = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 1 & 1 \\ -7 & 2 & -3 \end{bmatrix}$ the characteristic equation $|\mathbf{A} - \lambda\mathbf{I}| = 0$ reduces to $(\lambda-1)(\lambda-3)(\lambda+4) = 0$ so that the eigenroots are 1, 3 and −4. For $\lambda_* = 1$ the eigenvector, from the equation for $\mathbf{u}_*$,

$$\begin{aligned}
\mathbf{u}_* &= \left[\mathbf{I} - \begin{pmatrix} 1 & 2 & 0 \\ 2 & 0 & 1 \\ -7 & 2 & -4 \end{pmatrix}^{-} \begin{pmatrix} 1 & 2 & 0 \\ 2 & 0 & 1 \\ -7 & 2 & -4 \end{pmatrix}\right] \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \\
&= \left[\mathbf{I} + \frac{1}{4}\begin{pmatrix} 0 & -2 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 2 & 0 & 1 \\ -7 & 2 & -4 \end{pmatrix}\right] \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \\
&= \left[\mathbf{I} + \frac{1}{4}\begin{pmatrix} -4 & 0 & -2 \\ 0 & -4 & 1 \\ 0 & 0 & 0 \end{pmatrix}\right] \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & -\frac{1}{2} \\ 0 & 0 & \frac{1}{4} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}z_3 \\ \frac{1}{4}z_3 \\ z_3 \end{bmatrix} \quad \text{for any } z_3 \neq 0\,.
\end{aligned}$$

Similarly for $\lambda_* = 3$

$$\mathbf{u}_* = \left[\mathbf{I} - \begin{pmatrix} -1 & 2 & 0 \\ 2 & -2 & 1 \\ -7 & 2 & -3 \end{pmatrix}^{-} \begin{pmatrix} -1 & 2 & 0 \\ 2 & -2 & 1 \\ -7 & 2 & -3 \end{pmatrix}\right]\mathbf{z} = \left[\mathbf{I} - \tfrac{1}{2}\begin{pmatrix} -2 & -2 & 0 \\ -2 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} -1 & 2 & 0 \\ 2 & -2 & 1 \\ -7 & 2 & -3 \end{pmatrix}\right]\mathbf{z}$$

$$= \left[\mathbf{I} - \tfrac{1}{2}\begin{pmatrix} -2 & 0 & -2 \\ 0 & -2 & -1 \\ 0 & 0 & 0 \end{pmatrix}\right]\mathbf{z} = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & -\frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix}\mathbf{z} = \begin{bmatrix} -\,z_3 \\ -\frac{1}{2}z_3 \\ z_3 \end{bmatrix}.$$

The case of $\lambda_* = -4$ is left to the reader.

Properties of eigenroots

See the entry "Eigenroot".

Properties of eigenvectors

See the entry "Eigenvector".

SOME SUMMARIES

Orthogonal matrices

Any two of (i) $\mathbf{A}$ being square, (ii) $\mathbf{AA}' = \mathbf{I}$ and (iii) $\mathbf{A}'\mathbf{A} = \mathbf{I}$ imply the third; and define $\mathbf{A}$ as being orthogonal. Properties of orthogonal $\mathbf{A}$ include the following.

a. Rows (columns) are orthonormal.

b. $|\mathbf{A}| = \pm 1$.

c. λ being an eigenroot of $\mathbf{A}$ implies that $1/\lambda$ is also.

d. $\mathbf{AB}$ is orthogonal when $\mathbf{A}$ and $\mathbf{B}$ are.

Idempotent matrices

Idempotent $\mathbf{A}$ of order n has the following properties.

a. $\mathbf{A}^2 = \mathbf{A}$.

b. $\mathbf{A}$ is singular, unless $\mathbf{A} = \mathbf{I}$.

c. $\mathrm{r}(\mathbf{A}) = \mathrm{tr}(\mathbf{A})$.

d. $\mathbf{I} - \mathbf{A}$ is idempotent, with $\mathrm{r}(\mathbf{I} - \mathbf{A}) = \mathrm{n} - \mathrm{r}(\mathbf{A})$.

e. If $\mathbf{A}$ is also symmetric (but not $\mathbf{I}$) it is positive semi-definite, and can be expressed as $\mathbf{A} = \mathbf{LL}'$ for $\mathbf{L}'\mathbf{L} = \mathbf{I}$.

f. For idempotent $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{AB}$ is idempotent if $\mathbf{AB} = \mathbf{BA}$.

g. $r(\mathbf{A})$ eigenroots of $\mathbf{A}$ are 1.0, and $n - r(\mathbf{A})$ are 0.

h. There is a $\mathbf{U}$ such that

$$\mathbf{U}^{-1}\mathbf{AU} = \begin{bmatrix} \mathbf{I}_{r(A)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} .$$

i. $\mathbf{P} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ is idempotent, and is very useful in statistics.

Matrices aI + bJ

The matrix $a\mathbf{I} + b\mathbf{J}$ for $\mathbf{J} = \mathbf{1}\mathbf{1}'$ occurs in a number of analysis of variance situations in statistics. When of order n it has the following properties.

$$(a_1\mathbf{I} + b_1\mathbf{J})(a_2\mathbf{I} + b_2\mathbf{J}) = a_1a_2\mathbf{I} + (a_1b_2 + a_2b_1 + nb_1b_2)\mathbf{J} .$$

$$(\alpha\mathbf{I} + b\mathbf{J})^{-1} = \tfrac{1}{a}\left(\mathbf{I} - \frac{b}{a+nb}\mathbf{J}\right) .$$

$$|a\mathbf{I} + b\mathbf{J}| = a^{n-1}(a + nb) .$$

Eigenroots are a, $n - 1$ times, and $a + nb$ once.

Non-negative definite matrices

If a $\mathbf{A}$ is non-negative definite (n.n.d.):

$\mathbf{x}'\mathbf{Ax} \geq 0 \;\; \forall\; \mathbf{x} \neq 0$.

$\mathbf{A}$ is assumed symmetric because otherwise it can be replaced by $\tfrac{1}{2}(\mathbf{A} + \mathbf{A}')$.

$|\mathbf{A}| \geq 0$.

Diagonal elements of $\mathbf{A}$ are ≥ 0 .

Principal leading minors are ≥ 0 .

Eigenroots are ≥ 0 .

If $\mathbf{A}$ positive definite (p.d.) all the above ≥ 0 symbols become > 0 .

For real $\mathbf{X}$, $\mathbf{X}'\mathbf{X}$ is n.n.d. .

For real $\mathbf{X}$ of full column rank

$\mathbf{X'X}$ is p.d. .

$(\mathbf{X'X})^{-1}$ exists .

$\mathbf{XX'}$ has Moore-Penrose inverse $\mathbf{X(X'X)^{-2}X'}$.

Canonical and other forms

For any matrix $\mathbf{A}_{p\times q}$ of rank r:

(i) Equivalent canonical form:

$$\mathbf{PAQ} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{P} \text{ and } \mathbf{Q} \text{ non-singular} .$$

(ii) Similar canonical form:

$$\mathbf{AU} = \mathbf{UD}\{\lambda\} ,$$

where $\mathbf{D}\{\lambda\}$ is the diagonal matrix of eigenroots; and $\mathbf{U}$ is the matrix of corresponding eigenvectors. $\mathbf{U}^{-1}$ exists when the diagonability theorem is satisfied (see Eigenvector entry), and then

$$\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D}\{\lambda\} .$$

(iii) Singular-valued decomposition:

$$\mathbf{A} = \mathbf{L}\begin{bmatrix} \mathbf{\Delta}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{M}' ,$$

where $\mathbf{L}$ and $\mathbf{M}$ are each orthogonal, and $\mathbf{\Delta}_r = \sqrt{\Delta^2}$ where

$$\mathbf{L'AA'L} = \begin{bmatrix} \mathbf{\Delta}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \text{and} \quad \mathbf{M'A'AM} = \begin{bmatrix} \mathbf{\Delta}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

with $\mathbf{\Delta}^2$ being the diagonal matrix of the (positive) eigenroots of $\mathbf{A'A}$ (or, equivalently, of $\mathbf{AA'}$).

For symmetric $\mathbf{A}$ of order p and rank r:

(iv) Diagonal form:

$$\mathbf{PAP'} = \begin{bmatrix} \mathbf{D}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \text{ with } \mathbf{D}_r \text{ diagonal, order r} .$$

When $\mathbf{A}$ is n.n.d., elements of $\mathbf{D}_r$ are positive.

(v) Congruent canonical form:

$$\mathbf{RAR'} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \text{ for } \mathbf{R} \text{ possibly complex} .$$

When $\mathbf{A}$ is n.n.d., $\mathbf{R}$ is real.

(vi) Orthogonal similar canonical form:

$$\mathbf{U'AU} = \mathbf{D}\{\lambda\} \quad \text{with} \quad \mathbf{U} \text{ being orthogonal} \quad \text{and} \quad \mathbf{U}^{-1} = \mathbf{U'} .$$

(vii) Spectral decomposition:

$$\mathbf{A} = \Sigma_i \lambda_i \mathbf{u}_i \mathbf{u}_i'$$

for λ_i being an eigenroot and $\mathbf{u}_i$ its corresponding eigenvector.

SOLVING EQUATIONS BY ITERATION

Current computing facilities provide numerous methods of arithmetically solving equations which cannot be solved algebraically. Matrix notation permits succinct description of one of these methods.

For n equations in n unknowns represented by $\mathbf{x}$, let the equations be

$$\mathbf{f}(\mathbf{x}) = \mathbf{0} , \tag{1}$$

and define

$$\mathbf{G}(\mathbf{x}) = [g_{ij}(\mathbf{x})\} = \left\{\frac{\partial}{\partial x_j} f_i(\mathbf{x})\right\} \quad \text{for} \quad i,j = 1, \cdots, n . \tag{2}$$

Suppose $\mathbf{x}_r$ is an approximate solution for $\mathbf{x}$ to $\mathbf{f}(\mathbf{x}) = \mathbf{0}$. Then an improved approximation is $\mathbf{x}_{r+1}$ for

$$\mathbf{f}(\mathbf{x}_{r+1}) = \mathbf{f}(\mathbf{x}_r) + \mathbf{G}(\mathbf{x}_r)\boldsymbol{\Delta}_r \tag{3}$$

where

$$\boldsymbol{\Delta}_r = \mathbf{x}_{r+1} - \mathbf{x}_r . \tag{4}$$

Were $\mathbf{x}_{r+1}$ to be a solution to (1) then $\mathbf{f}(\mathbf{x}_{r+1})$ would be $\mathbf{0}$ and (3) would yield

$$\boldsymbol{\Delta}_r = -[\mathbf{G}(\mathbf{x}_r)]^{-1}\mathbf{f}(\mathbf{x}_r) \tag{5}$$

and with this (4) gives

$$\mathbf{x}_{r+1} = \boldsymbol{\Delta}_r + \mathbf{x}_r . \tag{6}$$

In this way (5) and (6) provide an iterative procedure for calculating a solution: for some initial value $\mathbf{x}_0$ use (5) to get $\boldsymbol{\Delta}_0$ and then (6) to get $\mathbf{x}_1$; and back to (5) to get $\boldsymbol{\Delta}_1$, and so on.

DIFFERENTIAL CALCULUS WITH MATRICES

A number of situations in statistics involve maximizing or minimizing a function: e.g., maximum likelihood estimation, least squares estimation, minimum variance procedures, minimizing less functions, and so on. In many cases differentiation of matrix expressions is involved, for which the following results are often useful.

Differentiating with respect to a scalar

Suppose elements of $\mathbf{A} = \{a_{ij}\}$ are functions of a scalar x. Then

$$\frac{\partial \mathbf{A}}{\partial x} = \left\{\frac{\partial a_{ij}}{\partial x}\right\},$$

$$\frac{\partial \mathbf{A}^{-1}}{\partial x} = -\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial x}\mathbf{A}^{-1},$$

$$\mathbf{A}\frac{\partial \mathbf{A}^{-}}{\partial x}\mathbf{A} = -\mathbf{A}\mathbf{A}^{-}\frac{\partial \mathbf{A}}{\partial x}\mathbf{A}^{-}\mathbf{A},$$

$$\mathbf{A}\frac{\partial (\mathbf{A}'\mathbf{A})^{-}}{\partial x}\mathbf{A}' = -\mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\frac{\partial (\mathbf{A}'\mathbf{A})}{\partial x}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}',$$

and for $\mathbf{A} = \mathbf{A}'$ and elements of $\mathbf{T}$ not involving x,

$$\mathbf{P} = \mathbf{T}(\mathbf{T}'\mathbf{A}\mathbf{T})^{-1}\mathbf{T}' \quad \text{has} \quad \frac{\partial \mathbf{P}}{\partial x} = -\mathbf{P}\frac{\partial \mathbf{A}}{\partial x}\mathbf{P}.$$

Differentiating with respect to elements of a vector

The basis of differentiating with respect to elements of $\mathbf{x}$ is defining what is meant by $\partial/\partial\mathbf{x}$. This is important because the definition determines the form of its various applications, and because not all writers use the same definition. Any presentation of this topic should therefore start with defining $\partial/\partial\mathbf{x}$.

A widely used convention is that for $\mathbf{x}$ being a column vector, $\partial/\partial\mathbf{x}$ is also:

$$\mathbf{x} = \left\{_c\; x_i\right\}_{i=1}^{n} \quad \text{defines} \quad \frac{\partial}{\partial \mathbf{x}} = \left\{_c\; \frac{\partial}{\partial x_i}\right\}_{i=1}^{n}.$$

Thus $\partial/\partial\mathbf{x}$ is a vector of differential operators. With this definition come the following results:

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{a}'\mathbf{x}) = \frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{a}) = \mathbf{a},$$

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{A}\mathbf{x}) = \mathbf{A}' \quad \text{and} \quad \frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{A}) = \mathbf{A},$$

$$\begin{aligned}\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}'\mathbf{A}\mathbf{x}) &= \mathbf{A}\mathbf{x} + \mathbf{A}'\mathbf{x} \text{ for } \mathbf{A} \text{ not symmetric}\\ &= 2\mathbf{A}\mathbf{x} \text{ for } \mathbf{A} \text{ symmetric}.\end{aligned}$$

Differentiating with respect to elements of a matrix

Again, the basic definition is important: for scalar θ and $\mathbf{X}_{p\times q}$

$$\frac{\partial \theta}{\partial \mathbf{X}} = \left\{_m\; \frac{\partial \theta}{\partial x_{ij}}\right\}_{i=1\;j=1}^{p\quad q}.$$

For $\mathbf{X}$ having functionally unrelated elements

$$\frac{\partial}{\partial \mathbf{X}}[\operatorname{tr}(\mathbf{X}\mathbf{A})] = \mathbf{A}'.$$

But for symmetric $\mathbf{X}$

$$\frac{\partial}{\partial \mathbf{X}}[\mathrm{tr}(\mathbf{XA})] = \mathbf{A} + \mathbf{A}' - \mathrm{diag}(\mathbf{A}) .$$

where diag(**A**) is the diagonal matrix of the diagonal elements of **A**. And, of course, these results also apply to tr(**AX**) = tr(**XA**).

Differentiating determinants

Let x_{ij} be the (i,j)th element of **X**, and $|\mathbf{X}_{ij}|$ its cofactor in $|\mathbf{X}|$. Then for **X** having functionally unrelated elements:

$$\frac{\partial|\mathbf{X}|}{\partial x_{ij}} = |\mathbf{X}_{ij}| ,$$

$$\frac{\partial|\mathbf{X}|}{\partial \mathbf{X}} = |\mathbf{X}|\,(\mathbf{X}^{-1})' ,$$

and

$$\frac{\partial}{\partial \mathbf{X}}\log|\mathbf{X}| = (\mathbf{X}^{-1})' .$$

For symmetric **X**, comparable results are

$$\frac{\partial|\mathbf{X}|}{\partial \mathbf{x}_{ij}} = (2 - \delta_{ij})\,|\mathbf{X}_{ij}| ,$$

where $\delta_{ij} = 0$ except when i = j and then $\delta_{ii} = 1$.

$$\frac{\partial|\mathbf{X}|}{\partial \mathbf{X}} = |\mathbf{X}|\,[2\mathbf{X}^{-1} - \mathrm{diag}(\mathbf{X}^{-1})]$$

and

$$\frac{\partial}{\partial \mathbf{X}}\log|\mathbf{X}| = 2\mathbf{X}^{-1} - \mathrm{diag}(\mathbf{X}^{-1}) .$$

Finally, for any non-singular **X**, symmetric or not,

$$\frac{\partial}{\partial \mathbf{y}}\log|\mathbf{X}| = \mathrm{tr}\left(\mathbf{X}^{-1}\frac{\partial \mathbf{X}}{\partial \mathbf{y}}\right).$$

Jacobians

When **y** is a vector of n differentiable functions of the n elements of **x**, such that the transformation of **x** to **y**, to be denoted $\mathbf{x}\rightarrow\mathbf{y}$, is 1-to-1, then the matrix

$$\mathbf{J}_{\mathbf{x}\rightarrow\mathbf{y}} = \left(\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\right)' = \left\{\frac{\partial x_i}{\partial y_j}\right\}_{i=1}^{n}\,_{j=1}^{n}$$

is the *Jacobian matrix* of $\mathbf{x}\rightarrow\mathbf{y}$. For example, if **y** = **Ax**,

$$\mathbf{J}_{\mathbf{x}\rightarrow\mathbf{y}} = \left[\frac{\partial(\mathbf{A}^{-1}\mathbf{y})}{\partial \mathbf{y}}\right]' = [(\mathbf{A}^{-1})']' = \mathbf{A}^{-1} .$$

$\|\mathbf{J}_{\mathbf{x}\rightarrow\mathbf{y}}\|$, the positive value of the determinant of $\mathbf{J}_{\mathbf{x}\rightarrow\mathbf{y}}$ is called the *Jacobian* of $\mathbf{x}\rightarrow\mathbf{y}$. It is needed when using $\mathbf{x}\rightarrow\mathbf{y}$ on an integral such as

$$\varphi = \int f(\mathbf{x})d\mathbf{x} ,$$

where f($\mathbf{x}$) is a scalar function of elements of $\mathbf{x}$. If $\mathbf{x}\rightarrow\mathbf{y}$ is $\mathbf{y} = g(\mathbf{x})$ then

$$\varphi = \int f(g^{-1}[\mathbf{y}])\, ||\, \mathbf{J}_{\mathbf{x}\rightarrow\mathbf{y}}\, ||\, d\mathbf{y}\ .$$

With the identity $||\,\mathbf{J}_{\mathbf{x}\rightarrow\mathbf{y}}\,|| \equiv 1/\,||\,\mathbf{J}_{\mathbf{y}\rightarrow\mathbf{x}}\,||$, with elements of $\mathbf{J}_{\mathbf{y}\rightarrow\mathbf{x}}$ sometimes being easier to derive than those of $\mathbf{J}_{\mathbf{x}\rightarrow\mathbf{y}}$, and when notation other than $\mathbf{x}$ and $\mathbf{y}$ is the context, confusion easily arises as to whether φ involves $\mathbf{J}_{\mathbf{x}\rightarrow\mathbf{y}}$ or $\mathbf{J}_{\mathbf{y}\rightarrow\mathbf{x}}$. Fortunately there is a mnemonic which clarifies the situation. Defining the transformation as old $\rightarrow$ new, one always uses $\mathbf{J}_{old\rightarrow new}$ abbreviated to $\mathbf{J}_{o\rightarrow n}$. In the latter the subscripts are always in the sequence "on", not "no". This always works.

VEC and VECH OPERATORS

Vectorizing a matrix can be done in various ways, the most useful of which is stacking the columns of a matrix one under the other. For $\mathbf{X}_{p\times q}$ the resulting column is denoted vec$\mathbf{X}$, a column of order pq. For example,

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 \\ a & b & c \end{bmatrix} \quad \text{gives} \quad \text{vec}\mathbf{X} = \begin{bmatrix} 1 \\ a \\ 2 \\ b \\ 3 \\ c \end{bmatrix}.$$

Three useful properties are

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\text{vec}\ \mathbf{B}$$

$$\text{tr}(\mathbf{AZ'BZC}) = (\text{vec}\mathbf{Z})'(\mathbf{A'C'} \otimes \mathbf{B})\text{vec}\ \mathbf{Z}$$

$$\text{tr}(\mathbf{AB}) = (\text{vec}\mathbf{A}')'\text{vec}\ \mathbf{B}\ .$$

The operator vech($\mathbf{X}$) is defined only for $\mathbf{X}$ being symmetric. It has the columns of $\mathbf{X}$, starting at the diagonal elements, stacked one under the other. For example

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & x & y \\ 3 & y & \alpha \end{bmatrix} \quad \text{has} \quad \text{vech}\mathbf{X} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ x \\ y \\ \alpha \end{bmatrix}.$$

Henderson and Searle (1979, 1981) give some history and numerous details.

A particular use of vec and vech is in calculating $\| \mathbf{J}_{X \to Y} \|$. It is the positive value of the determinant of

$$\mathbf{J}_{X \to Y} = \frac{\partial(\text{vec}\mathbf{X})}{\partial(\text{vec}\mathbf{Y})};$$

and if **X** and **Y** are both symmetric, vec is replaced by vech.

MATRICES HAVING COMPLEX NUMBERS AS ELEMENTS

Because statistics almost always deals with real numbers (e.g., data) and not complex numbers that involve $i = \sqrt{-1}$, most of this entry deals with real matrices, those having no complex numbers as elements. Nevertheless, since many texts do deal with matrices of complex numbers, a few basic definitions are given here.

In scalar arithmetic the complex number $a - ib$ is called the *complex conjugate* of $a + ib$, and the two numbers are a *conjugate pair*. Likewise with matrices: $\bar{\mathbf{M}} = \mathbf{A} - i\mathbf{B}$ is the *complex conjugate* of $\mathbf{M} = \mathbf{A} + i\mathbf{B}$, with **M** and $\bar{\mathbf{M}}$ being a *conjugate pair*. **M** is said to be *Hermitian* when $\bar{\mathbf{M}}' = \mathbf{M}$; and **M** is *unitary* if $\bar{\mathbf{M}}'\mathbf{M} = \mathbf{I}$. Thus being Hermitian is the complex counterpart of being symmetric, as is unitary of orthogonal.

SOME MATRIX USAGE IN STATISTICS

The development and description of statistical methodology benefits enormously from the use of matrices. The following examples briefly illustrate some of the widely used situations where matrix notation so efficiently encapsulates a multitude of results.

Means and variances

$\mathbf{x}$ being a vector of random variables with mean $\boldsymbol{\mu}$ implies $E(\mathbf{x}) = \boldsymbol{\mu}$, where E represents expectation. Then, because the i'th element of $\mathbf{x}$ has a variance, σ_i^2, and each pair of elements, the i'th and j'th say, have a covariance, σ_{ij}, these variances and covariances can be arrayed in a symmetric matrix, called the variance-covariance matrix. For example, for $\mathbf{x}$ of order 3,

$$\text{var}(\mathbf{x}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix} .$$

A more general expression is

$$\mathbf{V} = \text{var}(\mathbf{x}) = \text{E}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})' .$$

For a linear change of variables, from $\mathbf{x}$ to $\mathbf{y} = \mathbf{TX}$ the mean vector and the variance-covariance matrix are easily established as

$$\text{E}(\mathbf{y}) = \text{E}(\mathbf{Tx}) = \mathbf{T}\text{E}(\mathbf{x}) = \mathbf{T}\boldsymbol{\mu}$$

and

$$\begin{aligned} \text{var}(\mathbf{y}) = \text{var}(\mathbf{Tx}) &= \text{E}(\mathbf{TX} - \mathbf{T}\boldsymbol{\mu})(\mathbf{TX} - \mathbf{T}\boldsymbol{\mu})' \\ &= \text{E}\,\mathbf{T}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'\mathbf{T}' = \mathbf{TVT}' . \end{aligned}$$

Suppose $\mathbf{T}$ is a row vector, $\mathbf{t}'$. Then because a variance is never negative, $\text{var}(\mathbf{t}'\mathbf{y}) = \mathbf{t}'\mathbf{V}\mathbf{t} \geq 0$ and so $\mathbf{V}$ is n.n.d.

Correlation

A correlation matrix, $\mathbf{R}$ say, is a matrix with 1.0 as its diagonal elements and correlations $r_{ij} = \sigma_{ij}/\sqrt{\sigma_i^2\sigma_j^2}$ (for $i \neq j$) as its off-diagonal elements. Define $\mathbf{D}$ as the diagonal matrix of the σ_i^2 terms. Then $\mathbf{R} = \mathbf{D}^{-\frac{1}{2}}\mathbf{V}\mathbf{D}^{-\frac{1}{2}}$.

A frequently used form of $\mathbf{V}$ is one which has σ^2 for all diagonal elements (variances) and $\rho\sigma^2$ for all off-diagonal elements (covariances). Then

$$\mathbf{V} = \sigma^2\mathbf{R} \quad \text{for} \quad \mathbf{R} = (1 - \rho)\mathbf{I} + \rho\mathbf{J} ,$$

and for order k

$$|\mathbf{V}| = \sigma^{2k}(1 - \rho)^{k-1}(1 - \rho + k\rho) .$$

Since $\mathbf{V}$ is n.n.d. $|\mathbf{V}| \geq 0$ which implies $1 + (k - 1)\rho \geq 0$, i.e., $\rho \geq -1/(k - 1)$. This is a consequence which one would not be inclined to anticipate on assuming the same covariance, $\rho\sigma^2$, between each pair of variables.

Sums of squares and products

For a column vector $\mathbf{x}_j$, the j'th column of $\mathbf{X}$, the sum of squares of its elements $\mathbf{x}_{ij}$ is $\Sigma_i\mathbf{x}_{ij}^2 =$

$\mathbf{x}_j'\mathbf{x}_j$; and $\Sigma_i \mathbf{x}_{ij}\mathbf{x}_{ij'} = \mathbf{x}_j'\mathbf{x}_{j'}$. Thus $\mathbf{X}'\mathbf{X}$ is a matrix of these sums of squares and products.

For $\mathbf{x}_j$ having n elements, define $\mathbf{C}_n = \mathbf{I}_n - \bar{\mathbf{J}}_n$, the centering matrix of order n. Then $\mathbf{X}'\mathbf{C}_n\mathbf{X}$ has terms $\Sigma_i(\mathbf{x}_{ij} - \bar{\mathbf{x}}_{.j})^2$ in its diagonal and terms $\Sigma_i(\mathbf{x}_{ij} - \bar{\mathbf{x}}_{.j})(\mathbf{x}_{ij'} - \bar{\mathbf{x}}_{.j.})$ as its off-diagonal elements, with $\bar{\mathbf{x}}_{ij} = \mathbf{1}_n'\mathbf{x}_j/\mathrm{n}$. It is the matrix of sums of squares and products corrected for the mean.

The multivariate normal distribution

The density function of a normally distributed random variable x having mean μ and variance σ^2 is $[\exp - \frac{1}{2}(\mathrm{x} - \mu)^2/\sigma^2]/\sqrt{2\pi\sigma^2}$. The counterpart of this for a vector $\mathbf{x}$ of random variables distributed $\mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$, meaning that it has mean $\boldsymbol{\mu}$ and variance-covariance matrix $\mathbf{V}$, and having a multivariate normal distribution, is $[\exp - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{x} - \boldsymbol{\mu})]/\sqrt{|2\pi\mathbf{V}|}$. The moment generating function of linear combinations $\mathbf{Kx}$ of $\mathbf{x}$ is $\exp(\mathbf{t}'\mathbf{K}\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\mathbf{KVK}'\mathbf{t})$.

A very neat consequence of using matrices is the derivation of marginal and conditional distributions in the multivariate normal distribution $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$. It stems from partitioning $\mathbf{x}$, $\boldsymbol{\mu}$ and $\mathbf{V}$ as

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \quad \text{and} \quad \mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \text{with } \mathbf{V}_{21} = (\mathbf{V}_{12})' .$$

Then a marginal distribution is

$$\mathbf{x}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \mathbf{V}_{11})$$

and a conditional distribution is

$$\mathbf{x}_1|\mathbf{x}_2 \sim \mathcal{N}(\boldsymbol{\mu}_1 + \mathbf{V}_{12}\mathbf{V}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2),\ \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}] \ .$$

Details are available in Searle (1971, Section 2.4f).

Quadratic forms

Every sum of squares is a homogeneous second degree function of data. It can therefore be represented as a quadratic form $\mathbf{x}'\mathbf{Ax}$ for $\mathbf{x}$ being the vector of data and $\mathbf{A}$ being symmetric. A variety of properties pertaining to $\mathbf{x}'\mathbf{Ax}$ are then available for whatever sums of squares one is interested in. Some of these properties for $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{V})$ are as follows.

(i) $\mathrm{E}(\mathbf{x}'\mathbf{Ax}) = \mathrm{tr}(\mathbf{AV}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}$. (Normality is not needed for this result.)

(ii) $\mathrm{var}(\mathbf{x}'\mathbf{Ax}) = 2\mathrm{tr}(\mathbf{AV})^2 + 4\boldsymbol{\mu}'\mathbf{AVA}\boldsymbol{\mu}$.

(iii) $\mathbf{x'Ax}$ has a (non-control) chi-square distribution if and only if $\mathbf{AV}$ is idempotent .

(iv) $\mathbf{x'Ax}$ and $\mathbf{Lx}$ are stochastically independent if and only if $\mathbf{LVA} = \mathbf{0}$.

(v) $\mathbf{x'Ax}$ and $\mathbf{x'Bx}$ are stochastically independent if and only if $\mathbf{BVA} = \mathbf{0}$ or, equivalently, $\mathbf{AVB} = \mathbf{0}$.

Regression and linear models

There is an enormous literature on these topics, most of it using matrix algebra. Only a minute sampling of it is given here.

Consider a vector of data $\mathbf{y}$, modeled as having expected value $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ with $\mathbf{X}$ being known and $\boldsymbol{\beta}$ being a vector of unknown parameters. Defining $\boldsymbol{\epsilon}$ as $\mathbf{y} - \mathrm{E}(\mathbf{y})$, a vector of residuals leads to modeling $\mathbf{y}$ as $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Least squares estimation of $\boldsymbol{\beta}$ dictates minimizing $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ and taking the resulting value of $\boldsymbol{\beta}$, say $\hat{\boldsymbol{\beta}}$, as the estimator of $\boldsymbol{\beta}$. This leads to equations $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$. In regression $\mathbf{X}$ almost always has full column rank, so that $\mathbf{X}'\mathbf{X}$ is non-singular and hence $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. But with many more general linear models $(\mathbf{X}'\mathbf{X})^{-1}$ does not exist and a generalized inverse $(\mathbf{X}'\mathbf{X})^{-}$ has to be used. In that case there are many solutions for $\hat{\boldsymbol{\beta}}$ and to indicate this they can be denoted by $\boldsymbol{\beta}^o$. Thus $\boldsymbol{\beta}^o = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$.

Since $\boldsymbol{\beta}^o$ becomes $\hat{\boldsymbol{\beta}}$ when $(\mathbf{X}'\mathbf{X})^{-1}$ exists, properties of $\hat{\boldsymbol{\beta}}$ are included among those of $\boldsymbol{\beta}^o$, just a few of which are as follows.

(i) There are many solutions $\boldsymbol{\beta}^o$ but for each of them $\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta}^o = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$ is the same, because $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ is invariant to $(\mathbf{X}'\mathbf{X})^{-}$.

(ii) $\mathrm{E}(\boldsymbol{\beta}^o) \neq \boldsymbol{\beta}$, but $\mathrm{E}(\mathbf{X}\boldsymbol{\beta}^o) = \mathbf{X}\boldsymbol{\beta}$.

(iii) The residual sum of squares

$$\mathrm{SSE} = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$$

is invariant to $(\mathbf{X}'\mathbf{X})^{-}$. Because it can be expressed as $\mathrm{SSE} = \mathbf{y}'[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}']\mathbf{y}$ with the matrix being idempotent, the expected value of SSE for $\mathbf{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_N)$ is $\mathrm{E}(\mathrm{SSE}) = [\mathrm{N} - \mathrm{r}(\mathbf{X})]\sigma^2$. And SSE/σ^2 has a χ^2-distribution. Moreover, the sum of squares due to fitting the model is $\mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$; it too has a (non-central) χ^2-distribution, and it is stochastically independent of SSE.

And so on it goes. Readers whose appetite is whetted by this introduction to regression and linear models will find plenty of books and papers to satiate their hunger.

Shayle R. Searle

REFERENCES

Searle, S.R. (1982) *Matrix Algebra for Unbalanced Data.* John Wiley & Sons, New York.

Henderson, H.V. and Searle, S.R. (1979) Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics. *Canadian Journal of Statistics*, **7**, 65-81.

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

Searle, S.R. (1971) *Linear Models*, John Wiley & Sons, New York.