UNIVERSITY OF EDINBURGH
SCHOOL OF MATHEMATICS

Generalised Regression Models

GRM: Useful Matrix Results **Semester 1, 2022–2023**

---

1 Notation

In what follows a matrix will be denoted by a bold capital letter and its elements by the corresponding lower case letter with suffices. Thus the element in row i and column j of $\mathbf{A}$, or element (i,j), is a_{ij}. The transpose of $\mathbf{A}$ will be denoted by $\mathbf{A}^T$. Note, however, that some authors use a t or a prime rather than T and write $\mathbf{A}^t$ or $\mathbf{A}'$. A column vector will be denoted by a bold lower case letter, e.g. $\mathbf{x}$. The transpose $\mathbf{x}^T$ of $\mathbf{x}$ is a row vector. Note that if $\mathbf{x}$ is a column vector of order p with elements $x_1, x_2, \ldots, x_p$, then the product

$$\mathbf{x}^T\mathbf{x} = x_1^2 + x_2^2 + \ldots + x_p^2$$

is a scalar, whereas $\mathbf{x}\mathbf{x}^T$ is a $p \times p$ matrix. Any lower case letter that is not in bold will represent a scalar. The elements of all vectors and matrices are assumed to be real.

The *unit matrix*, which is a square matrix in which each diagonal element is 1 and each non-diagonal element is 0, will usually be denoted simply by $\mathbf{I}$. If, however, it is important to indicate the order of the matrix, say p, it will be written as $\mathbf{I}_p$. Similarly the *unit vector* of order p, each of whose elements is 1, will be denoted by either $\mathbf{1}$ or $\mathbf{1}_p$. Note that $\mathbf{1}_p^T\mathbf{1}_p = p$.

2 Trace of a Matrix

The *trace* of a square matrix $\mathbf{A} = [a_{ij}]$, denoted by $\operatorname{tr}\mathbf{A}$, is the sum of its diagonal elements. Thus, if $\mathbf{A}$ is a $p \times p$ matrix, then

$$\operatorname{tr}\mathbf{A} = \sum_{i=1}^{p} a_{ii}\,.$$

It may easily be verified that, if c is any scalar,

$$\begin{aligned}
\operatorname{tr}\mathbf{A}^T &= \operatorname{tr}\mathbf{A}\,,\\
\operatorname{tr}(\mathbf{A}+\mathbf{B}) &= \operatorname{tr}\mathbf{A} + \operatorname{tr}\mathbf{B}\,,\\
\operatorname{tr}(c\mathbf{A}) &= c\operatorname{tr}\mathbf{A}\,.
\end{aligned}$$

Suppose that $\mathbf{A}$ is a $p \times q$ matrix and that $\mathbf{B}$ is a $q \times p$ matrix, so that the products $\mathbf{AB}$ and $\mathbf{BA}$ both exist. Then it is easy to verify the following:

$$\operatorname{tr}(\mathbf{AB}) = \sum_{i=1}^{p}\sum_{j=1}^{q} a_{ij}b_{ji} = \operatorname{tr}(\mathbf{BA})\,, \tag{2.1}$$

$$\operatorname{tr}(\mathbf{ABC}) = \operatorname{tr}(\mathbf{BCA}) = \operatorname{tr}(\mathbf{CAB})\,, \tag{2.2}$$

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \operatorname{tr}(\mathbf{x}\mathbf{x}^T\mathbf{A}) = \operatorname{tr}(\mathbf{A}\mathbf{x}\mathbf{x}^T)\,. \tag{2.3}$$

3 Linear Equations and Transformations

For simultaneous linear equations in several variables, the notation of matrix algebra is particularly suitable. The set of p equations

$$a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{ip}x_p = c_i \quad (i = 1, \ldots, p)$$

in the p unknowns $x_1, x_2, \ldots, x_p$ can be written concisely as

$$\mathbf{Ax} = \mathbf{c}, \tag{3.1}$$

where $\mathbf{A}$ is the $p \times p$ matrix with elements a_{ij} and where $\mathbf{x}$ and $\mathbf{c}$ are column vectors with elements x_j and c_i respectively.

Suppose that $\mathbf{A}$ is non-singular, i.e. that its determinant $|\mathbf{A}|$ is non-zero. Then $\mathbf{A}$ has a (true) inverse $\mathbf{A}^{-1}$ satisfying

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

In this case the above set of equations has a unique solution which can be written as

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{c}.$$

If $\mathbf{A}$ is singular, i.e. if $|\mathbf{A}| = 0$, the equations do not have a unique solution. To obtain a solution, supposing that the equations are consistent, we can find a *generalized inverse* (or *pseudo-inverse*) of $\mathbf{A}$, which is a matrix $\mathbf{G}$ satisfying

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

Then $\mathbf{x} = \mathbf{Gc}$ is a solution of the equations (3.1), but is not unique. Note that some authors require $\mathbf{G}$ to satisfy one or more further conditions before it can be called a generalized inverse.

It is often necessary to transform linearly one set of variables, or coordinates, $x_1, \ldots, x_p$ into another set $y_1, \ldots, y_p$. The transformation giving the y's in terms of the x's and the inverse transformation giving the x's in terms of the y's may be represented by the equations

$$\mathbf{y} = \mathbf{Ax}, \quad \mathbf{x} = \mathbf{A}^{-1}\mathbf{y},$$

where $\mathbf{x}$ and $\mathbf{y}$ are the vectors with elements x_i and y_i respectively and where $\mathbf{A}$ is assumed to be non-singular.

4 Orthogonal Matrices and Transformations

In many cases both x's and y's represent rectangular, or orthogonal, coordinates. Both the transformation and the matrix defining it are then termed *orthogonal*. An orthogonal matrix $\mathbf{U}$ is a square matrix satisfying

$$\mathbf{UU}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I},$$

so that the rows and the columns of $\mathbf{U}$ each form a set of orthonormal vectors; thus

$$\mathbf{U}^T = \mathbf{U}^{-1}, \quad \mathbf{U} = (\mathbf{U}^T)^{-1}.$$

In view of this, an orthogonal transformation is represented by

$$\mathbf{y} = \mathbf{U}\mathbf{x}, \quad \mathbf{x} = \mathbf{U}^T\mathbf{y}.$$

Note that

$$\mathbf{y}^T\mathbf{y} = (\mathbf{x}^T\mathbf{U}^T)(\mathbf{U}\mathbf{x}) = \mathbf{x}^T(\mathbf{U}^T\mathbf{U})\mathbf{x} = \mathbf{x}^T\mathbf{I}\mathbf{x} = \mathbf{x}^T\mathbf{x}.$$

Considered geometrically, this expresses the fact that the distance of a point from the origin in a p-dimensional Euclidean space remains invariant under a change of coordinate axes.

For an orthogonal matrix $\mathbf{U}$ we have

$$|\mathbf{U}|^2 = |\mathbf{U}||\mathbf{U}^T| = |\mathbf{U}\mathbf{U}^T| = |\mathbf{I}| = 1,$$

and thus $|\mathbf{U}| = \pm 1$. The sign can always be made positive by changing, if necessary, the signs of all elements in one row or column of $\mathbf{U}$.

5 Linear Independence and Rank

A set of vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ is said to be *linearly independent* if there exists no set of scalars $c_1, \ldots, c_k$, not all zero, such that

$$c_1\mathbf{x}_1 + \ldots + c_k\mathbf{x}_k = \mathbf{0}.$$

A matrix (not necessarily square) is said to be of *rank r* if the maximum number of linearly independent rows (or, equivalently, the maximum number of linearly independent columns) is r. Alternatively and equivalently, the rank of $\mathbf{A}$ is defined to be r if every minor of order $r+1$ formed from $\mathbf{A}$ is zero and at least one minor of order r is not zero. If the rank of a square matrix $\mathbf{A}$ is less than its order, then $\mathbf{A}$ is singular, and conversely.

6 Eigenvalues and Eigenvectors of a Symmetric Matrix

Suppose that $\mathbf{A}$ is a symmetric $p \times p$ matrix. Then the *eigenvalues* (alternatively called *latent roots* or *characteristic roots*) of $\mathbf{A}$ are the values of λ (all real) satisfying the equation

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

which is a polynomial equation of degree p in λ. We may suppose that the eigenvalues are arranged in descending order, so that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$. Corresponding to the k-th value λ_k there is a vector $\mathbf{u}_k$, of order p, satisfying

$$\mathbf{A}\mathbf{u}_k = \lambda_k\mathbf{u}_k.$$

This is known as the k-th *eigenvector* of $\mathbf{A}$ (or the k-th *latent* or *characteristic* vector). If the λ_k are distinct, the vectors $\mathbf{u}_k$ are unique except that each may be multiplied by an arbitrary scalar. The vectors are also orthogonal to one another, i.e. are such that if $k \neq l$ then

$$\mathbf{u}_k^T\mathbf{u}_l = 0 \tag{6.1}$$

Even if the λ_k are not distinct, we may still choose the vectors $\mathbf{u}_k$ to be orthogonal to one another. We shall also suppose that they are *standardized*, i.e. that, for each value of k,

$$\mathbf{u}_k^T\mathbf{u}_k = 1. \tag{6.2}$$

Let $\mathbf{U}$ be the $p \times p$ matrix whose k-th column is $\mathbf{u}_k$. Then, in view of properties (6.1) and (6.2), $\mathbf{U}$ is an orthogonal matrix. Let Λ be the diagonal matrix of order p whose k-th diagonal element is λ_k. Then $\lambda_k \mathbf{u}_k$ is the k-th column of $\mathbf{U}\Lambda$. Since $\mathbf{A}\mathbf{u}_k$ is the k-th column of $\mathbf{AU}$, it follows that the equations satisfied by the $\mathbf{u}_k$ are equivalent to the matrix equation

$$\mathbf{AU} = \mathbf{U}\Lambda .$$

Post-multiplication of this by $\mathbf{U}^T$ gives

$$\mathbf{A} = \mathbf{U}\Lambda\mathbf{U}^T , \tag{6.3}$$

since $\mathbf{U}\mathbf{U}^T = \mathbf{I}$. It follows that the determinant of $\mathbf{A}$ is

$$\begin{aligned} |\mathbf{A}| &= |\mathbf{U}\Lambda\mathbf{U}^T| \\ &= |\mathbf{U}|\,|\Lambda|\,|\mathbf{U}^T| \\ &= |\Lambda| \\ &= \lambda_1\lambda_2\ldots\lambda_p , \end{aligned} \tag{6.4}$$

and that

$$\begin{aligned} \operatorname{tr}\mathbf{A} &= \operatorname{tr}(\mathbf{U}\Lambda\mathbf{U}^T) \\ &= \operatorname{tr}(\Lambda\mathbf{U}^T\mathbf{U}) \\ &= \operatorname{tr}\Lambda \\ &= \lambda_1 + \lambda_2 + \ldots + \lambda_p . \end{aligned} \tag{6.5}$$

Thus the determinant of $\mathbf{A}$ is the product of its eigenvalues, while the trace equals their sum.

The rank of $\mathbf{A}$ is equal to the number of non-zero eigenvalues. If none of these is zero, $\mathbf{A}$ is non-singular, and conversely. In that case $\mathbf{A}^{-1}$ exists and we have

$$\begin{aligned} \mathbf{A}^{-1} &= (\mathbf{U}\Lambda\mathbf{U}^T)^{-1} \\ &= (\mathbf{U}^T)^{-1}\Lambda^{-1}\mathbf{U}^{-1} \\ &= \mathbf{U}\Lambda^{-1}\mathbf{U}^T. \end{aligned} \tag{6.6}$$

The equations for $\mathbf{A}$ and $\mathbf{A}^{-1}$ in terms of Λ and $\mathbf{U}$ may alternatively be written as

$$\mathbf{A} = \sum_{k=1}^{p} \lambda_k \mathbf{u}_k \mathbf{u}_k^T \tag{6.7}$$

and

$$\mathbf{A}^{-1} = \sum_{k=1}^{p} \lambda_k^{-1} \mathbf{u}_k \mathbf{u}_k^T . \tag{6.8}$$

These are termed the *spectral decompositions* of $\mathbf{A}$ and $\mathbf{A}^{-1}$ respectively. Note that the eigenvalues of $\mathbf{A}^{-1}$ are $\lambda_1^{-1}, \ldots, \lambda_p^{-1}$ and that the eigenvectors of $\mathbf{A}^{-1}$ are the same as those of $\mathbf{A}$.

7 Quadratic Forms

The expression

$$\mathbf{x}^T\mathbf{A}\,\mathbf{x} = \sum_i \sum_j a_{ij} x_i x_j ,$$

where $\mathbf{A} = [a_{ij}]$ is symmetric, is known as a *quadratic form* in $\mathbf{x}$. Note that, if $i \neq j$, the coefficient of $x_i x_j$ is $2a_{ij}$, since $a_{ji} = a_{ij}$, whereas that of x_i^2 is a_{ii}.

Both the quadratic form and the matrix $\mathbf{A}$ are termed *positive definite* if $\mathbf{x}^T\mathbf{A}\,\mathbf{x} > 0$ for all non-null vectors $\mathbf{x}$. If $>$ is replaced by $\geq$, they are termed *non-negative definite*. They are called *positive semi-definite* if $\mathbf{x}^T\mathbf{A}\,\mathbf{x} \geq 0$ for all $\mathbf{x}$ and $\mathbf{x}^T\mathbf{A}\,\mathbf{x} = 0$ for some non-null $\mathbf{x}$.

With $\mathbf{U}$ and Λ as defined in Section 6, consider the orthogonal transformation given by

$$\mathbf{y} = \mathbf{U}^T\mathbf{x}, \; \mathbf{x} = \mathbf{U}\mathbf{y}.$$

Under this transformation we have

$$\begin{aligned} \mathbf{x}^T\mathbf{A}\mathbf{x} &= \mathbf{y}^T\mathbf{U}^\mathbf{T}(\mathbf{U}\Lambda\mathbf{U}^\mathbf{T})\mathbf{U}\mathbf{y} \\ &= \mathbf{y}^T(\mathbf{U}^T\mathbf{U})\Lambda(\mathbf{U}^T\mathbf{U})\,\mathbf{y} \\ &= \mathbf{y}^T\Lambda\mathbf{y} \qquad (7.1) \\ &= \sum_{k=1}^{p} \lambda_k y_k^2 \qquad (7.2) \end{aligned}$$

if $\mathbf{y}$ has elements $y_1, \ldots, y_p$. From this expression it is clear that a necessary and sufficient condition for $\mathbf{A}$ to be positive definite is that all its eigenvalues should be positive. For $\mathbf{A}$ to be non-negative definite the condition is that all eigenvalues should be non-negative. For $\mathbf{A}$ to be positive semi-definite all eigenvalues should be non-negative and at least one should be zero. If $\mathbf{A}$ is positive definite, then $\mathbf{A}^{-1}$ exists and is also positive definite.

8 Idempotent Matrices

A $p \times p$ matrix $\mathbf{H}$ is termed *idempotent* if $\mathbf{H} = \mathbf{H}^2$. Two examples with $p = 2$ are

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

If $\mathbf{H}$ has rank r, its eigenvalues consist of r unities and $p - r$ zeros, since if λ and $\mathbf{u}$ are an eigenvalue of $\mathbf{H}$ and the corresponding eigenvector then

$$\lambda\mathbf{u} = \mathbf{H}\mathbf{u} = \mathbf{H}^2\mathbf{u} = \mathbf{H}\lambda\mathbf{u} = \lambda^2\mathbf{u},$$

so that $\lambda^2 = \lambda$ and each eigenvalue is either 0 or 1. Putting the eigenvalues in descending order, we have $\lambda_1 = \lambda_2 = \ldots = \lambda_r = 1$ and $\lambda_{r+1} = \lambda_{r+2} = \ldots = \lambda_p = 0$. Hence, from (6.5),

$$\operatorname{tr}\mathbf{H} = \lambda_1 + \lambda_2 + \ldots + \lambda_p = r. \qquad (8.1)$$

Thus the rank of an idempotent matrix is the same as its trace, or the sum of its eigenvalues. In the trivial case where $r = p$, $\mathbf{H}$ is simply the unit matrix. If $\mathbf{H}$ is idempotent with rank r then $\mathbf{I} - \mathbf{H}$ is idempotent with rank $p - r$.

If $\mathbf{H}$ is both symmetric and idempotent, it may be expressed as

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T, \qquad (8.2)$$

where $\mathbf{X}$ is a $p \times r$ matrix of rank r. In this form it represents the projection operator on the column space of $\mathbf{X}$. If $\mathbf{u}_1, \ldots, \mathbf{u}_r$ are the eigenvectors corresponding to the unit eigenvalues of $\mathbf{H}$ and the $p \times r$ matrix $\mathbf{U}$ has columns $\mathbf{u}_1, \ldots, \mathbf{u}_r$ then, from equation (6.7),

$$\mathbf{H} = \sum_{k=1}^{r} \mathbf{u}_k \mathbf{u}_k^T = \mathbf{U}\mathbf{U}^T, \quad \mathbf{U}^T\mathbf{U} = \mathbf{I}_r . \tag{8.3}$$

Thus, if $\mathbf{x}$ is any p-vector and $\mathbf{z}$ denotes the r-vector $\mathbf{U}^T\mathbf{x}$, the quadratic form $\mathbf{x}^T\mathbf{H}\mathbf{x}$ becomes

$$\mathbf{x}^T\mathbf{H}\mathbf{x} = \mathbf{x}^T\mathbf{U}\mathbf{U}^T\mathbf{x} = \mathbf{z}^T\mathbf{z} = \sum_{k=1}^{r} z_k^2 , \tag{8.4}$$

a sum of squares of r variables.

If $r = 1$, $\mathbf{H}$ has the form $\mathbf{u}\mathbf{u}^T$, where $\mathbf{u}^T\mathbf{u} = 1$.

9 Partitioned Matrices

It is often convenient to represent a matrix in a partitioned form by the juxtaposition of two or more submatrices. An example of a partitioned matrix $\mathbf{A}$, not necessarily square, is

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

where $\mathbf{A}_{11}$ and $\mathbf{A}_{12}$ are submatrices having the same number of rows, $\mathbf{A}_{11}$ and $\mathbf{A}_{21}$ have the same number of columns, and so on. The transpose of $\mathbf{A}$ is

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

If both $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are square and symmetric and $\mathbf{A}_{12}^T = \mathbf{A}_{21}$, then $\mathbf{A}^T = \mathbf{A}$, so that $\mathbf{A}$ is square and symmetric.

Suppose that in the partitioned matrix

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

not necessarily square, the submatrices $\mathbf{B}_{11}$ and $\mathbf{B}_{12}$ have as many rows as $\mathbf{A}_{11}$ and $\mathbf{A}_{21}$ have columns, and that $\mathbf{B}_{21}$ and $\mathbf{B}_{22}$ have as many rows as $\mathbf{A}_{12}$ and $\mathbf{A}_{22}$ have columns. Then the product $\mathbf{AB}$ exists and is given by

$$\begin{aligned} \mathbf{AB} &= \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{bmatrix}. \end{aligned} \tag{9.1}$$

Thus the usual rule of matrix multiplication is applied with submatrices of $\mathbf{A}$ and $\mathbf{B}$ treated as elements.

Now suppose that $\mathbf{A}_{11}$ is a square $p \times p$ matrix and that $\mathbf{A}_{22}$ is a square $q \times q$ matrix (so that $\mathbf{A}_{12}$ is $p \times q$ and $\mathbf{A}_{21}$ is $q \times p$). Then, if $\mathbf{A}_{11}$ and $\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$ are non-singular, the inverse $\mathbf{A}^{-1}$ of the complete matrix $\mathbf{A}$ exists and there are convenient methods of finding it. Let the inverse of $\mathbf{A}$ be denoted by $\mathbf{B}$, where $\mathbf{B}$ is partitioned into submatrices $\mathbf{B}_{11}, \mathbf{B}_{12}, \mathbf{B}_{21}$ and $\mathbf{B}_{22}$ in exactly the same way as $\mathbf{A}$. Then the submatrices of $\mathbf{B} = \mathbf{A}^{-1}$ are given by

$$\begin{aligned}
\mathbf{B}_{22} &= (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}, \\
\mathbf{B}_{21} &= -\mathbf{B}_{22}\mathbf{A}_{21}\mathbf{A}_{11}^{-1}, \\
\mathbf{B}_{12} &= -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22}, \\
\mathbf{B}_{11} &= \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{22}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} \\
&= \mathbf{A}_{11}^{-1} - \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}_{21}\,.
\end{aligned} \tag{9.2}$$

These results may easily be verified by showing that $\mathbf{AB}$ equals $\mathbf{I}_{p+q}$.

The submatrices of $\mathbf{B} = \mathbf{A}^{-1}$ may alternatively be found by use of the formulae

$$\begin{aligned}
\mathbf{B}_{11} &= (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}, \\
\mathbf{B}_{12} &= -\mathbf{B}_{11}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}, \\
\mathbf{B}_{21} &= -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{B}_{11}, \\
\mathbf{B}_{22} &= \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{B}_{11}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\
&= \mathbf{A}_{22}^{-1} - \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{B}_{12}\,.
\end{aligned} \tag{9.3}$$

Which set of formulae should be chosen depends on the relative sizes of p and q and on which inverses are easiest to find.

Now consider the evaluation of the determinant of the partitioned matrix $\mathbf{A}$. The value of the determinant is unaltered if we subtract from the first p rows of $\mathbf{A}$, i.e. $[\mathbf{A}_{11}\ \mathbf{A}_{12}]$, the result of pre-multiplying the last q rows of $\mathbf{A}$, i.e. $[\mathbf{A}_{21}\ \mathbf{A}_{22}]$, by $\mathbf{A}_{12}\mathbf{A}_{22}^{-1}$. Thus

$$\begin{aligned}
|\mathbf{A}| &= \begin{vmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{vmatrix} \\
&= \begin{vmatrix} \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{O} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{vmatrix} \\
&= |\mathbf{A}_{22}| \left| \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} \right|.
\end{aligned} \tag{9.4}$$

Alternatively we can show in a similar manner that

$$|\mathbf{A}| = |\mathbf{A}_{11}| \left| \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \right|. \tag{9.5}$$

10 Vector Differentiation

Let $f = f(\mathbf{x})$ be a scalar function of the elements $x_1, \ldots, x_p$ of a vector $\mathbf{x}$ of order p. Then $df/d\mathbf{x}$ is the vector of order p whose i-th element is $\partial f/\partial x_i$.

Consider first differentiating the linear function

$$\mathbf{a}^T\mathbf{x} = \mathbf{x}^T\mathbf{a} = \sum_j a_j x_j\,,$$

in which $\mathbf{a}$ is a constant vector. The derivative of this with respect to x_i is a_i, which is element i of $\mathbf{a}$. Hence

$$d(\mathbf{a}^{\mathbf{T}}\mathbf{x})/d\mathbf{x} = \mathbf{a}\,. \tag{10.1}$$

Next consider the quadratic form

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \sum_j \sum_k a_{jk} x_j x_k \,,$$

where $\mathbf{A} = [a_{ij}]$ is a constant $p \times p$ symmetric matrix. Since $a_{ji} = a_{ij}$, the derivative of this expression with respect to x_i is $2\sum_j a_{ij} x_j$, which is element i of the vector $2\mathbf{Ax}$. Hence

$$d(\mathbf{x}^\mathbf{T}\mathbf{A}\mathbf{x})/d\mathbf{x} = 2\mathbf{A}\mathbf{x}\,. \tag{10.2}$$