Chapter 1. Matrix Algebra

1 Matrix and vector notation

Definition 1.1 *A* **matrix** *is a rectangular or square array of numbers of variables.*

We use uppercase boldface letters, and in this course all elements of matrices will be real numbers or variables representing real numbers. The notation $\boldsymbol{A} = (a_{ij})$ represents a matrix by means of a typical element, for example

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

Definition 1.2 *A* **vector** *is a matrix with a single column.*

We use lowercase boldface letters for column vectors, row vectors are expressed as transposes of column vectors, for example,

$$\boldsymbol{x}' = (x_1, x_2, x_3).$$

2 Rank

Definition 2.1 *A set of vectore* $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n$ *is said to be* **linearly dependent** *if scalars* $c_1, c_2, \cdots, c_n$ *(not all zero) can be found such that*

$$c_1\boldsymbol{a}_1 + c_2\boldsymbol{a}_2 + \cdots c_n\boldsymbol{a}_n = 0.$$

If no coefficients $c_1, c_2, \cdots, c_n$ *(not all zero) can be found that satisfy the above equation, the set of vectors* $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n$ *is said to be* **linearly independent***.*

Definition 2.2 *The* **rank** *of a matrix is the number of linearly independent rows (and columns) in the matrix.*

Theorem 2.1 *(i) If the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are conformable, then $rank(\boldsymbol{AB}) \leq rank(\boldsymbol{A})$ and $rank(\boldsymbol{AB}) \leq rank(\boldsymbol{B})$.*

(ii) If $\boldsymbol{B}$ and $\boldsymbol{C}$ are nonsingular, $rank(\boldsymbol{AB}) = rank(\boldsymbol{CA}) = rank(\boldsymbol{A})$.

(iii) For any matrix $\boldsymbol{A}$, $rank(\boldsymbol{A}'\boldsymbol{A}) = rank(\boldsymbol{AA}') = rank(\boldsymbol{A}') = rank(\boldsymbol{A})$.

PROOF: (i) All the columns of $\boldsymbol{AB}$ are linear combinations of the columns of $\boldsymbol{A}$. Consequently, the number of linearly independent columns of $\boldsymbol{AB}$ is less than or equal to the number of linearly independent columns of $\boldsymbol{A}$, and $rank(\boldsymbol{AB}) \leq rank(\boldsymbol{A})$. Similarly, all the rwos of $\boldsymbol{AB}$ are linear combinations of the rwos of $\boldsymbol{B}$, and therefore $rank(\boldsymbol{AB}) \leq rank(\boldsymbol{B})$.

(ii) If $\boldsymbol{B}$ is nonsingular, then there exists a matrix $\boldsymbol{B}^{-1}$ such that $\boldsymbol{BB}^{-1} = \boldsymbol{I}$. Then by part (i) we have

$$rank(\boldsymbol{A}) = rank(\boldsymbol{ABB}^{-1}) \leq rank(\boldsymbol{AB}) \leq rank(\boldsymbol{A}).$$

(iii) If $\boldsymbol{A}\boldsymbol{x} = 0$, $\boldsymbol{A}'\boldsymbol{A}\boldsymbol{x} = 0$. Hence, $rank(\boldsymbol{A}) \geq rank(\boldsymbol{A}'\boldsymbol{A})$. Conversely, if $\boldsymbol{A}'\boldsymbol{A}\boldsymbol{x} = 0$, $\boldsymbol{x}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{x} = (\boldsymbol{A}\boldsymbol{x})'(\boldsymbol{A}\boldsymbol{x}) = 0$, thus, $\boldsymbol{A}\boldsymbol{x} = 0$. Hence, $rank(\boldsymbol{A}'\boldsymbol{A}) \geq rank(\boldsymbol{A})$. So we have $rank(\boldsymbol{A}'\boldsymbol{A}) = rank(\boldsymbol{A})$. $\square$

3 Orthogonal Matrices

Definition 3.1 *A square matrix $\boldsymbol{A}$: $n \times n$ is said to be orthogonal if $\boldsymbol{A}\boldsymbol{A}' = \boldsymbol{I} = \boldsymbol{A}'\boldsymbol{A}$. For orthogonal matrices, we have*

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

(2) $|\boldsymbol{A}| = \pm 1$.

(3) Let $\delta_{ij} = 1$ for $i = j$ and 0 for $i \neq j$ denote the Kronecker symbol. The column vectors a_i and the row vectors $a_{(i)}$ of $\boldsymbol{A}$ satisfy the conditions

$$\boldsymbol{a}_i'\boldsymbol{a}_j = \delta_{ij}, \qquad \boldsymbol{a}_{(i)}\boldsymbol{a}_{(j)}' = \delta_{ij}.$$

(4) $\boldsymbol{AB}$ is orthogonal if $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal.

Theorem 3.1 *If the matrix $\boldsymbol{Q} : p \times p$ is orthogonal and if $\boldsymbol{A}$ is any $p \times p$ matrix, then*

(i) $|\boldsymbol{Q}| = +1$ *or* -1.

(ii) $|\boldsymbol{Q}'\boldsymbol{A}\boldsymbol{Q}| = |\boldsymbol{A}|$.

(iii) $-1 \leq c_{ij} \leq 1$, *where* c_{ij} *is any element of* $\boldsymbol{Q}$.

4 Eigenvalues and Eigenvectors

Definition 4.1 *If $\boldsymbol{A} : p \times p$ is a square matrix, then*

$$q(\lambda) = |\boldsymbol{A} - \lambda I|$$

is a pth order polynomial in λ. The p roots $\lambda_1, \cdots, \lambda_p$ of the characteristic equation $q(\lambda) = |\boldsymbol{A} - \lambda I| = 0$ are called eigenvalues or characteristic roots of $\boldsymbol{A}$.

The eigenvalues possibly may be complex numbers. Since $|A - \lambda_i I| = 0$, $\boldsymbol{A} - \lambda_i I$ is a singular matrix. Hence, there exists a nonzero vector $\boldsymbol{\gamma}_i \neq 0$ satisfying $(\boldsymbol{A} - \lambda_i I)\boldsymbol{\gamma}_i = 0$, that is,

$$\boldsymbol{A}\boldsymbol{\gamma}_i = \lambda_i \boldsymbol{\gamma}_i,$$

$\boldsymbol{\gamma}_i$ is called the (right) eigenvectors of $\boldsymbol{A}$ for the eigenvalue λ_i. If λ_i is complex, then $\boldsymbol{\gamma}_i$ may have complex components. An eigenvector $\boldsymbol{\gamma}$ with real components is called standardized if $\boldsymbol{\gamma}'\boldsymbol{\gamma} = 1$.

Theorem 4.1 *If $\boldsymbol{A}$ is any $n \times n$ matrix with eigenvalues $\lambda_1, \cdots, \lambda_n$, then*

(i) $|\boldsymbol{A}| = \prod_{i=1}^{n} \lambda_i$.

(ii) $trace(\boldsymbol{A}) = \sum_{i=1}^{n} \lambda_i$.

Theorem 4.2 *Let $\boldsymbol{A}$ be an $n \times n$ symmetric matrix.*

(i) The eigenvalues are all real.

(ii) The eigenvectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$ of $\boldsymbol{A}$ are mutually orthogonal; that is, $\boldsymbol{x}_i' \boldsymbol{x}_j = 0$ for $i \neq j$.

Theorem 4.3 *If λ is an eigenvalue of the square matrix $\boldsymbol{A}$ with corresponding eigenvector $\boldsymbol{x}$, then for certain functions $g(\boldsymbol{A})$, an eigenvalue is given by $g(\lambda)$ and $\boldsymbol{x}$ is the corresponding eigenvector of $g(\boldsymbol{A})$ as well as $\boldsymbol{A}$.*

Example 4.1

$$\begin{aligned}(\boldsymbol{A}^3 + 4\boldsymbol{A}^2 - 3\boldsymbol{A} + 5I)\boldsymbol{x} &= \boldsymbol{A}^3\boldsymbol{x} + 4\boldsymbol{A}^2\boldsymbol{x} - 3\boldsymbol{A}\boldsymbol{x} + 5\boldsymbol{x} \\ &= (\lambda^3 + 4\lambda^2 - 3\lambda + 5)\boldsymbol{x}.\end{aligned}$$

Thus $\lambda^3 + 4\lambda^2 - 3\lambda + 5$ *is an eigenvalue of* $\boldsymbol{A}^3 + 4\boldsymbol{A}^2 - 3\boldsymbol{A} + 5I$*, and* $\boldsymbol{x}$ *is the corresponding eigenvector.*

5 Decomposition of Matrices

Theorem 5.1 *(Spectral decomposition theorem) If* $\boldsymbol{A}$ *is an* $n \times n$ *symmetric matrix with eigenvalues* $\lambda_1, \lambda_2, \cdots, \lambda_n$ *and normalized eigenvectors* $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$*, then* $\boldsymbol{A}$ *can be expressed as*

$$\boldsymbol{A} = \boldsymbol{C}\boldsymbol{D}\boldsymbol{C}' = \sum_{i=1}^{n} \lambda_i \boldsymbol{x}_i \boldsymbol{x}_i',$$

where $\boldsymbol{D} = diag(\lambda_1, \cdots, \lambda_n)$ *and* $\boldsymbol{C}$ *is the orthogonal matrix* $\boldsymbol{C} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)$*.*

PROOF: See Rencher and Schaalje (2008, pp.51-52) □

Theorem 5.2 *(Singular Value Decomposition) Suppose that $\boldsymbol{A}$ is an $n \times p$ matrix of rank r. Then there exists two orthogonal matrices $\boldsymbol{P}_{n\times n}$ and $\boldsymbol{Q}_{p\times p}$ such that*

$$\boldsymbol{A} = \boldsymbol{P} \begin{pmatrix} \boldsymbol{A}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix} \boldsymbol{Q}',$$

where $\boldsymbol{A}_r = diag(\lambda_1, \cdots, \lambda_r)$, $\lambda_i > 0$, $i = 1, 2, \cdots, r$, and $\lambda_1^2, \cdots, \lambda_r^2$ are non-zero eigenvalues of $\boldsymbol{A}'\boldsymbol{A}$.

6 Positive Definite and Positive Semidefinite Matrices

Definition 6.1 *If a symmetric matrix $\boldsymbol{A}$ has the property $\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y} > 0$ for all possible $\boldsymbol{y}$ except $\boldsymbol{y} = 0$, $\boldsymbol{A}$ is said to be a positive definite matrix and it is denoted by $\boldsymbol{A} > 0$. Similarly, if $\boldsymbol{y}'\boldsymbol{A}\boldsymbol{y} \geq 0$ for all $\boldsymbol{y}$ except $\boldsymbol{y} = 0$, $\boldsymbol{A}$ is said to be positive semidefinite and it is denoted by $\boldsymbol{A} \geq 0$.*

Theorem 6.1 *Let $\boldsymbol{A}$ be $n \times n$ with eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$.*

(i) If $\boldsymbol{A}$ is positive definite, then $\lambda_i > 0$ for $i = 1, 2, \cdots, n$.

(ii) If $\boldsymbol{A}$ is positive semidefinite, then $\lambda_i \geq 0$ for $i = 1, 2, \cdots, n$. The number of nonzero eigenvalues is the rank of $\boldsymbol{A}$.

If a matrix $\boldsymbol{A}$ is positive definite, we can find a square root matrix $\boldsymbol{A}^{1/2}$,

$$\boldsymbol{A}^{1/2} = \boldsymbol{C}\boldsymbol{D}^{1/2}\boldsymbol{C}',$$

where $\boldsymbol{D}^{1/2} = diag(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_n})$. The matrix $\boldsymbol{A}^{1/2}$ is symmetric and has the property

$$\boldsymbol{A}^{1/2}\boldsymbol{A}^{1/2} = (\boldsymbol{A}^{1/2})^2 = \boldsymbol{A}.$$

7 Idempotent Matrices

Definition 7.1 *A square matrix $\boldsymbol{A}$ is called idempotent if it satisfies*

$$\boldsymbol{A}^2 = \boldsymbol{A}.$$

An idempotent matrix $\boldsymbol{A}$ is called an orthogonal projector if $\boldsymbol{A} = \boldsymbol{A}'$. Otherwise, $\boldsymbol{A}$ is called an oblique projector.

Theorem 7.1 *Let $\boldsymbol{A}$ be an $n \times n$ idempotent matrix with rank$(\boldsymbol{A}) = r \leq n$. Then we have:*

- *(i)* *The eigenvalues of $\boldsymbol{A}$ are 1 or 0.*
- *(ii)* $tr(A) = rank(A) = r$.
- *(iii)* *If $\boldsymbol{A}$ is of full rank n, then $\boldsymbol{A} = \boldsymbol{I}_n$.*
- *(iv)* *If $\boldsymbol{A}$ and $\boldsymbol{B}$ are idempotent and if $\boldsymbol{AB} = \boldsymbol{BA}$, then $\boldsymbol{AB}$ is also idempotent.*
- *(v)* *If $\boldsymbol{A}$ is idempotent and $\boldsymbol{P}$ is orthogonal, then $\boldsymbol{PAP}'$ is also idempotent.*
- *(vi)* *If $\boldsymbol{A}$ is idempotent, then $\boldsymbol{I} - \boldsymbol{A}$ is idempotent and*

$$\boldsymbol{A}(\boldsymbol{I} - \boldsymbol{A}) = (\boldsymbol{I} - \boldsymbol{A})\boldsymbol{A} = 0.$$

8 Generalized inverse

Definition 8.1 *Let $\boldsymbol{A}$ be an $n \times p$ matrix. Then a $p \times n$ matrix $\boldsymbol{A}^{-}$ is said to be a generalized inverse of $\boldsymbol{A}$ if*

$$\boldsymbol{A}\boldsymbol{A}^{-}\boldsymbol{A} = \boldsymbol{A}.$$

holds.

Note when $\boldsymbol{A}$ is nonsingular, $\boldsymbol{A}^{-}$ is unique and $\boldsymbol{A}^{-} = \boldsymbol{A}^{-1}$. A generalized inverse is also called a conditional inverse. If $\boldsymbol{A}$ is $n \times p$, any generalized inverse $\boldsymbol{A}^{-}$ is $p \times n$.

Every matrix, whether square or rectangular, has a generalized inverse. For example,

$$\boldsymbol{x} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}.$$

Then, $\boldsymbol{x}_1^- = (1, 0, 0, 0)$, $\boldsymbol{x}_2^- = (0, 1/2, 0, 0)$, $\boldsymbol{x}_3^- = (0, 0, 1/3, 0)$, and $\boldsymbol{x}_4^- = (0, 0, 0, 1/4)$ are all generalized inverses of $\boldsymbol{x}$.

Example 8.1

$$\boldsymbol{A} = \begin{pmatrix} 2 & 2 & 3 \\ 1 & 0 & 1 \\ 3 & 2 & 4 \end{pmatrix}.$$

One $\boldsymbol{A}^-$ is:

$$\boldsymbol{A}^- = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Theorem 8.1 *Suppose $\boldsymbol{A}$ is $n \times p$ of rank r and that $\boldsymbol{A}$ is partitioned as*

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

where $\boldsymbol{A}_{11}$ is $r \times r$ of rank r. Then a generalized inverse of $\boldsymbol{A}$ is given by

$$\boldsymbol{A}^{-} = \begin{pmatrix} \boldsymbol{A}_{11}^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix},$$

where the three $\boldsymbol{O}$ matrices are of appropriate sizes so that $\boldsymbol{A}^{-}$ is $p \times n$.

PROOF: By multiplication of partitioned matrices, we obtain

$$\boldsymbol{A}\boldsymbol{A}^{-}\boldsymbol{A} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{O} \\ \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{O} \end{pmatrix} \boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12} \end{pmatrix}.$$

To show $\boldsymbol{A}_21\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12} = \boldsymbol{A}_{22}$, we multiply $\boldsymbol{A}$ by

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{O} \\ -\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{I} \end{pmatrix}$$

and we have

$$\boldsymbol{BA} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{O} & \boldsymbol{A}_{22} - \boldsymbol{A}_21\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12} \end{pmatrix}$$

Since $\boldsymbol{B}$ is nonsingular, $rank(\boldsymbol{BA}) = rank(\boldsymbol{A}) = r$. Since $\boldsymbol{A}_{11}$ is nonsingular, there exists a matrix $\boldsymbol{D}$ such that

$$\begin{pmatrix} \boldsymbol{A}_{12} \\ \boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}_{11} \\ \boldsymbol{O} \end{pmatrix} \boldsymbol{D}.$$

Hence $\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12} = \boldsymbol{O}$ and the theorem is proved. $\square$

Corollary 8.1 *Suppose* $\boldsymbol{A}$ *is* $n \times p$ *of rank* r *and that* $\boldsymbol{A}$ *is partitioned as*

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

where $\boldsymbol{A}_{22}$ *is* $r \times r$ *of rank* r*. Then a generalized inverse of* $\boldsymbol{A}$ *is given by*

$$\boldsymbol{A}^{-} = \begin{pmatrix} \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{A}_{22}^{-1} \end{pmatrix},$$

where the three $\boldsymbol{O}$ *matrices are of appropriate sizes so that* $\boldsymbol{A}^{-}$ *is* $p \times n$*.*

The nonsingular submatrix need not be in the $\boldsymbol{A}_{11}$ or $\boldsymbol{A}_{12}$ positions, as in the above theorem and corollary. In this case, we can do row and column operations on $\boldsymbol{A}$ to get a $r \times r$ leading minor. That is, there exists nonsingular matrix $\boldsymbol{P}_{m\times m}$ and $\boldsymbol{Q}_{n\times n}$ such that

$$\boldsymbol{RAC} = \boldsymbol{B} = \begin{pmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{pmatrix},$$

where $\boldsymbol{B}_{11}$ is $r \times r$ with $rank(\boldsymbol{B}_{11}) = r$. It can be shown that

$$\boldsymbol{G} = \boldsymbol{C} \begin{pmatrix} \boldsymbol{B}_{11}^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix} \boldsymbol{R}$$

is a generalized inverse of $\boldsymbol{A}$.

More general, we can find many generalized inverse of $\boldsymbol{A}$ by using the theorem of singular value decomposition, there exists nonsingular matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ such that

$$\boldsymbol{PAQ} = \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix}.$$

Then

$$\boldsymbol{G} = \boldsymbol{Q} \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{G}_{12} \\ \boldsymbol{G}_{21} & \boldsymbol{G}_{22} \end{pmatrix} \boldsymbol{P}$$

is a generalized inverse of $\boldsymbol{A}$, where $\boldsymbol{G}_{12}, \boldsymbol{G}_{21}$ and $\boldsymbol{G}_{22}$ are arbitrary with conformable sizes.

To see this:

$$\boldsymbol{A} = \boldsymbol{P}^{-1} \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix} \boldsymbol{Q}^{-1}$$

$$\begin{aligned} \boldsymbol{AGA} &= \boldsymbol{P}^{-1} \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix} \boldsymbol{Q}^{-1}\boldsymbol{Q} \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{G}_{12} \\ \boldsymbol{G}_{21} & \boldsymbol{G}_{22} \end{pmatrix} \boldsymbol{P}\boldsymbol{P}^{-1} \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix} \boldsymbol{Q}^{-1} \\ &= \boldsymbol{P}^{-1} \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{G}_{12} \\ \boldsymbol{G}_{21} & \boldsymbol{G}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix} \boldsymbol{Q}^{-1} \\ &= \boldsymbol{P}^{-1} \begin{pmatrix} \boldsymbol{I}_r & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix} \boldsymbol{Q}^{-1} \\ &= \boldsymbol{A} \end{aligned}$$

For the above result, we can see that

(1) There exist infinitely many generalized inverse for any singular matrix.

(2) $rank(\boldsymbol{A}^{-}) \geq rank(\boldsymbol{A})$ (we can choose $\boldsymbol{A}^{-}$ with $rank(\boldsymbol{A}^{-}) = rank(\boldsymbol{A})$).

(3) If $\boldsymbol{A}$ is symmetric, $\boldsymbol{A}^{-}$ needs not be symmetric.

In practice, we do not need to carry out row and column operations to get a $r \times r$ nonsingular leading minor in order to find a generalized inverse. A general algorithm for finding a generalized inverse $\boldsymbol{A}^-$ for any $n \times p$ matrix $\boldsymbol{A}$ of rank r (Searle 1982, P.218) is as follows.

1. Find any nonsingular $r \times r$ submatrix $\boldsymbol{C}$. It is not necessary that the elements of $\boldsymbol{C}$ occupy adjacent rows and columns in $\boldsymbol{A}$.

2. Find $\boldsymbol{C}^{-1}$ and $(\boldsymbol{C}^{-1})'$.

3. Replace the elements of $\boldsymbol{C}$ by the elements of $(\boldsymbol{C}^{-1})'$.

4. Replace all other elements in $\boldsymbol{A}$ by zeros.

5. Transpose the resulting matrix.

Some properties of generalized inverses are given in the following theorem.

Lemma 8.1 *(i) For any matrix* $\boldsymbol{A}$ *we have* $\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{O}$ *if and only if* $\boldsymbol{A} = \boldsymbol{O}$*.*

(ii) Let $\boldsymbol{X} \neq \boldsymbol{O}$ *be an* $m \times n$ *matrix and* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *be* $n \times n$ *matrices,* $\boldsymbol{A}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{B}\boldsymbol{X}'\boldsymbol{X}$ *if and only if* $\boldsymbol{A}\boldsymbol{X}' = \boldsymbol{B}\boldsymbol{X}'$*.*

PROOF: (i) If $\boldsymbol{A} = \boldsymbol{O}$, we have $\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{O}$. Conversely, if $\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{O}$, let $\boldsymbol{A} = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n)$ be the columnwise partition, then

$$\boldsymbol{A}'\boldsymbol{A} = (\boldsymbol{a}_i'\boldsymbol{a}_j) = \boldsymbol{O},$$

so that $\boldsymbol{a}_i = 0$ and $\boldsymbol{A} = \boldsymbol{O}$.

(ii) If $\boldsymbol{A}\boldsymbol{X}' = \boldsymbol{B}\boldsymbol{X}'$, we have $\boldsymbol{A}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{B}\boldsymbol{X}'\boldsymbol{X}$. Conversely, if $\boldsymbol{A}\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{B}\boldsymbol{X}'\boldsymbol{X}$, we have

$$(\boldsymbol{A} - \boldsymbol{B})\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{A}' - \boldsymbol{B}') = \boldsymbol{O}.$$

So $(\boldsymbol{A} - \boldsymbol{B})\boldsymbol{X}' = \boldsymbol{O}$, i.e., $\boldsymbol{A}\boldsymbol{X}' = \boldsymbol{B}\boldsymbol{X}'$. $\square$

Theorem 8.2 *Let $\boldsymbol{A}$ be $n \times p$ of rank r, let $\boldsymbol{A}^-$ be any generalized inverse of $\boldsymbol{A}$, and let $(\boldsymbol{A}'\boldsymbol{A})^-$ be any generalized inverse of $\boldsymbol{A}'\boldsymbol{A}$. Then:*

(i) *rank($\boldsymbol{A}^-\boldsymbol{A}$)=rank($\boldsymbol{A}\boldsymbol{A}^-$)=rank($\boldsymbol{A}$)=$r$.*

(ii) *$(\boldsymbol{A}^-)'$ is a generalized inverse of $\boldsymbol{A}'$; that is, $(\boldsymbol{A}')^- = (\boldsymbol{A}^-)'$.*

(iii) *$\boldsymbol{A} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^-\boldsymbol{A}'\boldsymbol{A}$ and $\boldsymbol{A}' = \boldsymbol{A}'\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^-\boldsymbol{A}'$.*

(iv) *$(\boldsymbol{A}'\boldsymbol{A})^-\boldsymbol{A}'$ is a generalized inverse of $\boldsymbol{A}$; that is, $\boldsymbol{A}^- = (\boldsymbol{A}'\boldsymbol{A})^-\boldsymbol{A}'$.*

(v) *$\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^-\boldsymbol{A}'$ is symmetric, has rank=r, and is invariant to the choice of $(\boldsymbol{A}'\boldsymbol{A})^-$; that is, $\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^-\boldsymbol{A}'$ remains the same, no matter what value of $(\boldsymbol{A}'\boldsymbol{A})^-$ is used.*

PROOF: (i) The result follows from

$$rank(\boldsymbol{A}) = rank(\boldsymbol{A}\boldsymbol{A}^{-}\boldsymbol{A}) \leq rank(\boldsymbol{A}^{-}\boldsymbol{A}) \leq rank(\boldsymbol{A}).$$

(ii) Since $\boldsymbol{A}\boldsymbol{A}^{-}\boldsymbol{A} = \boldsymbol{A}$, and $\boldsymbol{A}'(\boldsymbol{A}^{-})'\boldsymbol{A}' = \boldsymbol{A}'$, hence, $(\boldsymbol{A}^{-})'$ is a generalized inverse of $\boldsymbol{A}'$.

(iii) Let $\boldsymbol{B} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-}\boldsymbol{A}'\boldsymbol{A} - \boldsymbol{A}$, we can verify that

$$\boldsymbol{B}'\boldsymbol{B} = (\boldsymbol{A}'\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-}\boldsymbol{A}' - \boldsymbol{A}')(\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-}\boldsymbol{A}'\boldsymbol{A} - \boldsymbol{A}) = \boldsymbol{O}$$

by the definition of a generalized inverse.

(iv) It follows from (iii).

(v) Since $(\boldsymbol{A}'\boldsymbol{A})^{-}\boldsymbol{A}'$ is a generalized inverse of $\boldsymbol{A}$, and (i) $rank(\boldsymbol{A}\boldsymbol{A}^{-}) = r$, so $rank(\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-}\boldsymbol{A}') = r$.

Let $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ be any two generalized inverse of $\boldsymbol{A}'\boldsymbol{A}$, from (iii) we have

$$\boldsymbol{A}\boldsymbol{G}_1\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{A}\boldsymbol{G}_2\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{A}.$$

By using the lemma, we have $\boldsymbol{A}\boldsymbol{G}_1\boldsymbol{A}' = \boldsymbol{A}\boldsymbol{G}_2\boldsymbol{A}'$, which means that $\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-}\boldsymbol{A}'$ is invariant to the choice of $(\boldsymbol{A}'\boldsymbol{A})^{-}$. $\square$

Definition 8.2 *(Moore-Penrose inverse) A matrix* $\boldsymbol{A}^+$ *satisfying the following conditions is called the Moore-Penrose inverse of* $\boldsymbol{A}$*:*

(i) $\boldsymbol{A}\boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{A}$,

(ii) $\boldsymbol{A}^+\boldsymbol{A}\boldsymbol{A}^+ = \boldsymbol{A}^+$,

(iii) $(\boldsymbol{A}^+\boldsymbol{A})' = \boldsymbol{A}^+\boldsymbol{A}$,

(iv) $(\boldsymbol{A}\boldsymbol{A}^+)' = \boldsymbol{A}\boldsymbol{A}^+$.

$\boldsymbol{A}^+$ *is unique.*

PROOF: Proof for the uniqueness of Moore-Penrose inverse.

Suppose that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are both $\boldsymbol{A}^+$. From the definition of Moore-Penrose inverse, we have

$$\begin{aligned}\boldsymbol{X} &= \boldsymbol{XAX} = \boldsymbol{X}(\boldsymbol{AX})' = \boldsymbol{XX}'\boldsymbol{A}' = \boldsymbol{XX}'(\boldsymbol{AYA})' = \boldsymbol{X}(\boldsymbol{AX})'(\boldsymbol{AY})' \\ &= (\boldsymbol{XAX})\boldsymbol{AY} = \boldsymbol{XAY} = (\boldsymbol{XA})'\boldsymbol{YAY} = \boldsymbol{A}'\boldsymbol{X}'\boldsymbol{A}'\boldsymbol{Y}'\boldsymbol{Y} \\ &= \boldsymbol{A}'\boldsymbol{Y}'\boldsymbol{Y} = (\boldsymbol{YA})'\boldsymbol{Y} = \boldsymbol{YAY} = \boldsymbol{Y}.\end{aligned}$$

This completes the proof. □

$\boldsymbol{A}^+$ is a special case of $\boldsymbol{A}^-$. It possesses all of the properties that $\boldsymbol{A}^-$ has. In addition, it is unique and possesses the following properties.

Theorem 8.3 *Let $\boldsymbol{A}$ be an $m \times n$-matrix. Then*

(i) $\boldsymbol{A}$ *regular, then* $\boldsymbol{A}^{+} = \boldsymbol{A}^{-1}$.

(ii) $(\boldsymbol{A}^{+})^{+} = A$.

(iii) $(\boldsymbol{A}^{+})' = (\boldsymbol{A}')^{+}$.

(iv) $rank(\boldsymbol{A}) = rank(\boldsymbol{A}^{+}) = rank(\boldsymbol{A}^{+}\boldsymbol{A}) = rank(\boldsymbol{A}\boldsymbol{A}^{+})$.

(v) *If* $rank(A) = m$, $\boldsymbol{A}^{+} = \boldsymbol{A}'(\boldsymbol{A}\boldsymbol{A}')^{-1}$ *and* $\boldsymbol{A}\boldsymbol{A}^{+} = \boldsymbol{I}_m$.

(vi) *If* $rank(A) = n$, $\boldsymbol{A}^{+} = (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$ *and* $\boldsymbol{A}^{+}\boldsymbol{A} = \boldsymbol{I}_n$.

(vii) *If* $\boldsymbol{P}$ *is an* $m \times m$ *orthogonal matrix,* $\boldsymbol{Q}$ *is an* $n \times n$ *orthogonal matrix,* $(\boldsymbol{P}\boldsymbol{A}\boldsymbol{Q})^{+} = \boldsymbol{Q}^{-1}\boldsymbol{A}^{+}\boldsymbol{P}^{-1}$.

(viii) $(\boldsymbol{A}'\boldsymbol{A})^{+} = \boldsymbol{A}^{+}(\boldsymbol{A}')^{+}$ *and* $(\boldsymbol{A}\boldsymbol{A}')^{+} = (\boldsymbol{A}')^{+}\boldsymbol{A}^{+}$.

(ix) $\boldsymbol{A}^{+} = (\boldsymbol{A}'\boldsymbol{A})^{+}\boldsymbol{A}' = \boldsymbol{A}'(\boldsymbol{A}\boldsymbol{A}')^{+}$.

Theorem 8.4 *Suppose that $\boldsymbol{A}$ can be expressed as in the singular value decomposition theorem, then*

$$\boldsymbol{A}^{+} = \boldsymbol{Q} \begin{pmatrix} \boldsymbol{A}_r^{-1} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{pmatrix} \boldsymbol{P}'.$$

For any matrix $\boldsymbol{A}$, $\boldsymbol{A}^{+}$ is unique.

PROOF: It can be easily verified that the $\boldsymbol{A}^{+}$ defined in the theorem satisfies all the equations given in the definition of Moore-Penrose inverse. □

Theorem 8.5 *Let* $\boldsymbol{A}$ *be an* $n \times n$ *symmetric matrix,* $\boldsymbol{a} \in R(\boldsymbol{A})$*,* $\boldsymbol{b} \in R(\boldsymbol{A})$*, and assume* $1 + \boldsymbol{b}'\boldsymbol{A}^{+}\boldsymbol{a} \neq 0$*. Then*

$$(\boldsymbol{A} + \boldsymbol{a}\boldsymbol{b}')^{+} = \boldsymbol{A}^{+} - \frac{\boldsymbol{A}^{+}\boldsymbol{a}\boldsymbol{b}'\boldsymbol{A}^{+}}{1 + \boldsymbol{b}'\boldsymbol{A}^{+}\boldsymbol{a}}.$$

9 Generalized Inverse and Systems of Equations

Theorem 9.1 *The system of equations $\boldsymbol{Ax} = \boldsymbol{c}$ has a solution (consistent) if and only if for any generalized inverse $\boldsymbol{A}^-$ of $\boldsymbol{A}$*

$$\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{c} = \boldsymbol{c}.$$

PROOF: (sufficiency) If the system is consistent, then $\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{c} = \boldsymbol{c}$. Since $\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A} = \boldsymbol{A}$, we have

$$\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x}.$$

Substituting $\boldsymbol{Ax} = \boldsymbol{c}$ on both sides, we have

$$\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{c} = \boldsymbol{c}.$$

(necessity) If $\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{c} = \boldsymbol{c}$, the system is consistent. Suppose $\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{c} = \boldsymbol{c}$, a solution exists, namely $\boldsymbol{x} = \boldsymbol{A}^-\boldsymbol{c}$. □

Theorem 9.2 *If the system of equations $\boldsymbol{Ax} = \boldsymbol{c}$ is consistent, then all possible solutions can be obtained in the following two ways.*

(a) Use all possible values of $\boldsymbol{A}^{-}$ in $\boldsymbol{x} = \boldsymbol{A}^{-}\boldsymbol{c}$.

(b) Use a specific $\boldsymbol{A}^{-}$ in $\boldsymbol{x} = \boldsymbol{A}^{-}\boldsymbol{c} + (\boldsymbol{I} - \boldsymbol{A}^{-}\boldsymbol{A})\boldsymbol{z}$, and use all possible values of the arbitrary vector $\boldsymbol{z}$. In particularly, we write $\boldsymbol{x} = \boldsymbol{A}^{+}\boldsymbol{c} + (\boldsymbol{I} - \boldsymbol{A}^{+}\boldsymbol{A})\boldsymbol{z}$. Among all the solutions, $\boldsymbol{x}_0 = \boldsymbol{A}^{+}\boldsymbol{c}$ has the smallest length.

PROOF: Part (a): see Searle (1982, p.238). Part II of (b):

$$\begin{aligned}
\|\boldsymbol{x}\|^2 &= [\boldsymbol{A}^+\boldsymbol{c} + (\boldsymbol{I} - \boldsymbol{A}^+\boldsymbol{A})\boldsymbol{z}]'[\boldsymbol{A}^+\boldsymbol{c} + (\boldsymbol{I} - \boldsymbol{A}^+\boldsymbol{A})\boldsymbol{z}] \\
&= \|\boldsymbol{x}\|^2 + \boldsymbol{z}'(\boldsymbol{I} - \boldsymbol{A}^+\boldsymbol{A})^2\boldsymbol{z} + 2\boldsymbol{c}'(\boldsymbol{A}^+)'(\boldsymbol{I} - \boldsymbol{A}^+\boldsymbol{A})z \\
&= \|\boldsymbol{x}\|^2 + \boldsymbol{z}'(\boldsymbol{I} - \boldsymbol{A}^+\boldsymbol{A})^2\boldsymbol{z} \\
&\geq \|\boldsymbol{x}\|^2.
\end{aligned}$$

This is because $(\boldsymbol{A}^+)'(\boldsymbol{I} - \boldsymbol{A}^+\boldsymbol{A}) = \boldsymbol{O}$, and $\boldsymbol{z}'(\boldsymbol{I} - \boldsymbol{A}^+\boldsymbol{A})^2\boldsymbol{z} \geq 0$ for arbitrary $\boldsymbol{z}$. The results follows. □

10 Derivatives of Linear Functions and Quadratic Forms

Let $u = f(\boldsymbol{x})$ be a function of the variable $x_1, x_2, \cdots, x_p$ in $\boldsymbol{x} = (x_1, x_2, \cdots, x_p)'$, and let $\partial u/\partial x_1, \cdots, \partial u/\partial x_p$ be the partial derivatives. We define $\partial u/\partial \boldsymbol{x}$ as

$$\frac{\partial u}{\partial \boldsymbol{x}} = \begin{pmatrix} \frac{\partial u}{\partial x_1} \\ \frac{\partial u}{\partial x_2} \\ \vdots \\ \frac{\partial u}{\partial x_p} \end{pmatrix}.$$

Theorem 10.1 *Let $u = \boldsymbol{a}'\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{a}$, where $\boldsymbol{a}' = (a_1, a_2, \cdots, a_p)$ is a vector of constants. Then*

$$\frac{\partial u}{\partial \boldsymbol{x}} = \frac{\partial(\boldsymbol{a}'\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\partial(\boldsymbol{x}'\boldsymbol{a})}{\partial \boldsymbol{x}} = \boldsymbol{a}.$$

Theorem 10.2 *Let* $u = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$, *where* $\boldsymbol{A}$ *is a symmetric matrix of constants. Then*

$$\frac{\partial u}{\partial \boldsymbol{x}} = \frac{\partial(\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = 2\boldsymbol{A}\boldsymbol{x}.$$