Chapter 1

Linear and Matrix Algebra

This chapter summarizes some important results of linear and matrix algebra that are instrumental in deriving many statistical results in subsequent chapters. Our emphasis is given to special matrices and their properties. Although the coverage of these mathematical topics is rather brief, it is self-contained. Readers may also consult other linear and matrix algebra textbooks for more detailed discussions; see e.g., Anton (1981), Basilevsky (1983), Graybill (1969), and Noble and Daniel (1977).

1.1 Basic Notations

A *matrix* is an array of numbers. In what follows, a matrix is denoted by an upper-case alphabet in boldface (e.g., $\boldsymbol{A}$), and its (i,j) th element (the element at the i th row and j th column) is denoted by the corresponding lower-case alphabet with subscripts ij (e.g., a_{ij}). Specifically, a $m \times n$ matrix $\boldsymbol{A}$ contains m rows and n columns and can be expressed as

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}.$$

An $n \times 1$ $(1 \times n)$ matrix is an n-dimensional column (row) *vector*. Every vector will be denoted by a lower-case alphabet in boldface (e.g., $\boldsymbol{z}$), and its i th element is denoted by the corresponding lower-case alphabet with subscript i (e.g., z_i). An 1×1 matrix is just a *scalar*. For a matrix $\boldsymbol{A}$, its i th column is denoted as $\boldsymbol{a}_i$.

A matrix is *square* if its number of rows equals the number of columns. A matrix is said to be *diagonal* if its off-diagonal elements (i.e., a_{ij}, $i \neq j$) are all zeros and at least one of its diagonal elements is non-zero, i.e., $a_{ii} \neq 0$ for some $i = 1, \dots, n$. A diagonal matrix whose diagonal elements are all ones is an *identity* matrix, denoted as $\boldsymbol{I}$; we also write the $n \times n$ identity matrix as $\boldsymbol{I}_n$. A matrix $\boldsymbol{A}$ is said to be *lower* (*upper*) *triangular* if $a_{ij} = 0$ for $i < (>)\, j$. We let $\mathbf{0}$ denote the matrix whose elements are all zeros.

For a vector-valued function $\boldsymbol{f} : \mathbb{R}^m \to \mathbb{R}^n$, $\nabla_{\boldsymbol{\theta}} \boldsymbol{f}(\boldsymbol{\theta})$ is the $m \times n$ matrix of the first-order derivatives of $\boldsymbol{f}$ with respect to the elements of $\boldsymbol{\theta}$:

$$\nabla_{\boldsymbol{\theta}} \boldsymbol{f}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial f_1(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial f_2(\boldsymbol{\theta})}{\partial \theta_1} & \cdots & \frac{\partial f_n(\boldsymbol{\theta})}{\partial \theta_1} \\ \frac{\partial f_1(\boldsymbol{\theta})}{\partial \theta_2} & \frac{\partial f_2(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial f_n(\boldsymbol{\theta})}{\partial \theta_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1(\boldsymbol{\theta})}{\partial \theta_m} & \frac{\partial f_2(\boldsymbol{\theta})}{\partial \theta_m} & \cdots & \frac{\partial f_n(\boldsymbol{\theta})}{\partial \theta_m} \end{bmatrix}.$$

When $n = 1$, $\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta})$ is the (column) *gradient* vector of $f(\boldsymbol{\theta})$. The $m \times m$ *Hessian* matrix of the second-order derivatives of the real-valued function $f(\boldsymbol{\theta})$ is

$$\nabla^2_{\boldsymbol{\theta}} f(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}(\nabla_{\boldsymbol{\theta}} f(\boldsymbol{\theta})) = \begin{bmatrix} \frac{\partial^2 f(\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_1} & \frac{\partial^2 f(\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 f(\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_m} \\ \frac{\partial^2 f(\boldsymbol{\theta})}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 f(\boldsymbol{\theta})}{\partial \theta_2 \partial \theta_2} & \cdots & \frac{\partial^2 f(\boldsymbol{\theta})}{\partial \theta_2 \partial \theta_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\boldsymbol{\theta})}{\partial \theta_m \partial \theta_1} & \frac{\partial^2 f(\boldsymbol{\theta})}{\partial \theta_m \partial \theta_2} & \cdots & \frac{\partial^2 f(\boldsymbol{\theta})}{\partial \theta_m \partial \theta_m} \end{bmatrix}.$$

1.2 Matrix Operations

Two matrices are said to be of the same size if they have the same number of rows and same number of columns. Matrix equality is defined for two matrices of the same size. Given two $m \times n$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, $\boldsymbol{A} = \boldsymbol{B}$ if $a_{ij} = b_{ij}$ for every i, j. The *transpose* of an $m \times n$ matrix $\boldsymbol{A}$, denoted as $\boldsymbol{A}'$, is the $n \times m$ matrix whose $(i, j)\,\mathrm{th}$ element is the $(j, i)\,\mathrm{th}$ element of $\boldsymbol{A}$. The transpose of a column vector is a row vector; the transpose of a scalar is just the scalar itself. A matrix $\boldsymbol{A}$ is said to be *symmetric* if $\boldsymbol{A} = \boldsymbol{A}'$, i.e., $a_{ij} = a_{ji}$ for all i, j. Clearly, a diagonal matrix is symmetric, but a triangular matrix is not.

Matrix addition is also defined for two matrices of the same size. Given two $m \times n$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, their sum, $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$, is the $m \times n$ matrix with the $(i, j)\,\mathrm{th}$ element $c_{ij} = a_{ij} + b_{ij}$. Note that matrix addition, if defined, is commutative:

$$\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A},$$

and associative:

$$\boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C}) = (\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C}.$$

Also, $\boldsymbol{A} + \mathbf{0} = \boldsymbol{A}$.

The scalar multiplication of the scalar c and matrix $\boldsymbol{A}$ is the matrix $c\boldsymbol{A}$ whose (i,j) th element is ca_{ij}. Clearly, $c\boldsymbol{A} = \boldsymbol{A}c$, and $-\boldsymbol{A} = -1\times\boldsymbol{A}$. Thus, $\boldsymbol{A}+(-\boldsymbol{A}) = \boldsymbol{A}-\boldsymbol{A} = \boldsymbol{0}$. Given two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, the matrix multiplication $\boldsymbol{AB}$ is defined only when the number of columns of $\boldsymbol{A}$ is the same as the number of rows of $\boldsymbol{B}$. Specifically, when $\boldsymbol{A}$ is $m\times n$ and $\boldsymbol{B}$ is $n\times p$, their product, $\boldsymbol{C} = \boldsymbol{AB}$, is the $m\times p$ matrix whose (i,j) th element is

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

Matrix multiplication is not commutative, i.e., $\boldsymbol{AB} \neq \boldsymbol{BA}$; in fact, when $\boldsymbol{AB}$ is defined, $\boldsymbol{BA}$ need not be defined. On the other hand, matrix multiplication is associative:

$$\boldsymbol{A}(\boldsymbol{BC}) = (\boldsymbol{AB})\boldsymbol{C},$$

and distributive with respect to matrix addition:

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

It is easy to verify that $(\boldsymbol{AB})' = \boldsymbol{B}'\boldsymbol{A}'$. For an $m\times n$ matrix $\boldsymbol{A}$, $\boldsymbol{I}_m\boldsymbol{A} = \boldsymbol{A}\boldsymbol{I}_n = \boldsymbol{A}$.

The *inner product* of two d-dimensional vectors $\boldsymbol{y}$ and $\boldsymbol{z}$ is the scalar

$$\boldsymbol{y}'\boldsymbol{z} = \sum_{i=1}^{d} y_i z_i.$$

If $\boldsymbol{y}$ is m-dimensional and $\boldsymbol{z}$ is n-dimensional, their *outer product* is the matrix $\boldsymbol{y}\boldsymbol{z}'$ whose (i,j) th element is $y_i z_j$. In particular,

$$\boldsymbol{z}'\boldsymbol{z} = \sum_{i=1}^{d} z_i^2,$$

which is non-negative and induces the standard *Euclidean norm* of $\boldsymbol{z}$ as $\|\boldsymbol{z}\| = (\boldsymbol{z}'\boldsymbol{z})^{1/2}$. The vector with Euclidean norm zero must be a zero vector; the vector with Euclidean norm one is referred to as a *unit vector*. For example,

$$(1\ 0\ 0), \qquad \left(0\ \frac{1}{2}\ \frac{\sqrt{3}}{2}\right), \qquad \left(\frac{1}{\sqrt{2}}\ \frac{1}{\sqrt{3}}\ \frac{1}{\sqrt{6}}\right),$$

are all unit vectors. A vector whose i th element is one and the remaining elements are all zero is called the i th Cartesian unit vector.

Let θ denote the angle between $\boldsymbol{y}$ and $\boldsymbol{z}$. By the law of cosine,

$$\|\boldsymbol{y}-\boldsymbol{z}\|^2 = \|\boldsymbol{y}\|^2 + \|\boldsymbol{z}\|^2 - 2\|\boldsymbol{y}\|\,\|\boldsymbol{z}\|\cos\theta,$$

where the left-hand side is $\|\boldsymbol{y}\|^2 + \|\boldsymbol{z}\|^2 - 2\boldsymbol{y}'\boldsymbol{z}$. Thus, the inner product of $\boldsymbol{y}$ and $\boldsymbol{z}$ can be expressed as

$$\boldsymbol{y}'\boldsymbol{z} = \|\boldsymbol{y}\|\|\boldsymbol{z}\| \cos\theta.$$

When $\theta = \pi/2$, $\cos\theta = 0$ so that $\boldsymbol{y}'\boldsymbol{z} = 0$. In this case, we say that $\boldsymbol{y}$ and $\boldsymbol{z}$ are *orthogonal* to each other. A square matrix $\boldsymbol{A}$ is said to be *orthogonal* if $\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}' = \boldsymbol{I}$. Hence, each column (row) vector of an orthogonal matrix is a unit vector and orthogonal to all remaining column (row) vectors. When $\boldsymbol{y} = c\boldsymbol{z}$ for some $c \neq 0$, $\theta = 0$ or π, and $\boldsymbol{y}$ and $\boldsymbol{z}$ are said to be *linearly dependent*.

As $-1 \leq \cos\theta \leq 1$, we immediately obtain the so-called *Cauchy-Schwarz inequality*.

Lemma 1.1 (Cauchy-Schwarz) *For two d-dimensional vectors $\boldsymbol{y}$ and $\boldsymbol{z}$,*

$$|\boldsymbol{y}'\boldsymbol{z}| \leq \|\boldsymbol{y}\|\|\boldsymbol{z}\|,$$

where the equality holds when $\boldsymbol{y}$ and $\boldsymbol{z}$ are linearly dependent.

It follows from the Cauchy-Schwarz inequality that

$$\begin{aligned}\|\boldsymbol{y}+\boldsymbol{z}\|^2 &= \|\boldsymbol{y}\|^2 + \|\boldsymbol{z}\|^2 + 2\boldsymbol{y}'\boldsymbol{z} \\ &\leq \|\boldsymbol{y}\|^2 + \|\boldsymbol{z}\|^2 + 2\|\boldsymbol{y}\|\|\boldsymbol{z}\| \\ &= (\|\boldsymbol{y}\| + \|\boldsymbol{z}\|)^2.\end{aligned}$$

This leads to the following *triangle inequality*.

Lemma 1.2 *For two d-dimensional vectors $\boldsymbol{y}$ and $\boldsymbol{z}$,*

$$\|\boldsymbol{y}+\boldsymbol{z}\| \leq \|\boldsymbol{y}\| + \|\boldsymbol{z}\|,$$

where the equality holds when $\boldsymbol{y} = c\boldsymbol{z}$ for some $c > 0$.

When $\boldsymbol{y}$ and $\boldsymbol{z}$ are orthogonal,

$$\|\boldsymbol{y}+\boldsymbol{z}\|^2 = \|\boldsymbol{y}\|^2 + \|\boldsymbol{z}\|^2,$$

which is the celebrated *Pythagoras theorem*.

A special type of matrix multiplication, known as the *Kronecker product*, is defined for matrices without size restrictions. Specifically, the Kronecker product of two matrices $\boldsymbol{A}$ ($m \times n$) and $\boldsymbol{B}$ ($p \times q$) is the $mp \times nq$ matrix:

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \cdots & a_{2n}\boldsymbol{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & a_{m2}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{bmatrix}.$$

The Kronecker product is not commutative:

$$\boldsymbol{A} \otimes \boldsymbol{B} \neq \boldsymbol{B} \otimes \boldsymbol{A},$$

but it is associative:

$$(\boldsymbol{A} \otimes \boldsymbol{B}) \otimes \boldsymbol{C} = \boldsymbol{A} \otimes (\boldsymbol{B} \otimes \boldsymbol{C}).$$

It also obeys the distributive law:

$$\boldsymbol{A} \otimes (\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{A} \otimes \boldsymbol{B} + \boldsymbol{A} \otimes \boldsymbol{C}.$$

It can be verified that

$$(\boldsymbol{A} \otimes \boldsymbol{B})' = \boldsymbol{A}' \otimes \boldsymbol{B}'.$$

Consider now differentiation involving vectors and matrices. Let $\boldsymbol{a}$ and $\boldsymbol{\theta}$ be tow d-dimensional vectors. We have

$$\nabla_{\boldsymbol{\theta}} (\boldsymbol{a}'\boldsymbol{\theta}) = \boldsymbol{a}.$$

For a symmetric matrix $\boldsymbol{A}$,

$$\nabla_{\boldsymbol{\theta}} (\boldsymbol{\theta}'\boldsymbol{A}\boldsymbol{\theta}) = 2\boldsymbol{A}\boldsymbol{\theta}, \qquad \nabla^2_{\boldsymbol{\theta}} (\boldsymbol{\theta}'\boldsymbol{A}\boldsymbol{\theta}) = 2\boldsymbol{A}.$$

1.3 Matrix Determinant and Trace

Given a square matrix $\boldsymbol{A}$, let $\boldsymbol{A}_{ij}$ denote the sub-matrix obtained from $\boldsymbol{A}$ by deleting its i th row and j th column. The *determinant* of $\boldsymbol{A}$ is

$$\det(\boldsymbol{A}) = \sum_{i=1}^{m} (-1)^{i+j} a_{ij} \det(\boldsymbol{A}_{ij}),$$

for any $j = 1, \ldots, n$, where $(-1)^{i+j} \det(\boldsymbol{A}_{ij})$ is called the *cofactor* of a_{ij}. This definition is based on the cofactor expansion along the j th column. Equivalently, the determinant can also be defined using the cofactor expansion along the i th row:

$$\det(\boldsymbol{A}) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(\boldsymbol{A}_{ij}),$$

for any $i = 1, \ldots, m$. The determinant of a scalar is the scalar itself; the determinant of a 2×2 matrix $\boldsymbol{A}$ is simply $a_{11}a_{22} - a_{12}a_{21}$. A square matrix with non-zero determinant is said to be *nonsingular*; otherwise, it is *singular*.

Clearly, $\det(\boldsymbol{A}) = \det(\boldsymbol{A}')$. From the definition of determinant, it is straightforward to see that for a scalar c and an $n \times n$ matrix $\boldsymbol{A}$,

$$\det(c\boldsymbol{A}) = c^n \det(\boldsymbol{A}),$$

and that for a square matrix with a column (or row) of zeros, its determinant must be zero. Also, the determinant of a diagonal or triangular matrix is simply the product of all the diagonal elements. It can also be shown that the determinant of the product of two square matrices of the same size is the product of their determinants:

$$\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B}) = \det(\boldsymbol{BA}).$$

Also, for an $m \times m$ matrix $\boldsymbol{A}$ and a $p \times p$ matrix $\boldsymbol{B}$,

$$\det(\boldsymbol{A} \otimes \boldsymbol{B}) = \det(\boldsymbol{A})^m \det(\boldsymbol{B})^p.$$

If $\boldsymbol{A}$ is an orthogonal matrix, we know $\boldsymbol{AA}' = \boldsymbol{I}$ so that

$$\det(\boldsymbol{I}) = \det(\boldsymbol{AA}') = [\det(\boldsymbol{A})]^2.$$

As the determinant of the identity matrix is one, the determinant of an orthogonal matrix must be either 1 or -1.

The *trace* of a square matrix is the sum of its diagonal elements; i.e., $\mathrm{trace}(\boldsymbol{A}) = \sum_i a_{ii}$. For example, $\mathrm{trace}(\boldsymbol{I}_n) = n$. Clearly, $\mathrm{trace}(\boldsymbol{A}) = \mathrm{trace}(\boldsymbol{A}')$. The trace function has the linear property:

$$\mathrm{trace}(c\boldsymbol{A} + d\boldsymbol{B}) = c\,\mathrm{trace}(\boldsymbol{A}) + d\,\mathrm{trace}(\boldsymbol{B}),$$

where c and d are scalars. It can also be shown that

$$\mathrm{trace}(\boldsymbol{AB}) = \mathrm{trace}(\boldsymbol{BA}),$$

provided that both $\boldsymbol{AB}$ and $\boldsymbol{BA}$ are defined. For two square matrices $\boldsymbol{A}$ and$\boldsymbol{B}$,

$$\mathrm{trace}(\boldsymbol{A} \otimes \boldsymbol{B}) = \mathrm{trace}(\boldsymbol{A})\,\mathrm{trace}(\boldsymbol{B}).$$

1.4 Matrix Inverse

A nonsingular matrix $\boldsymbol{A}$ possesses a unique *inverse* $\boldsymbol{A}^{-1}$ in the sense that $\boldsymbol{AA}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$. A singular matrix cannot be inverted, however. Thus, saying that a matrix is invertible is equivalent to saying that it is nonsingular.

Given an invertible matrix $\boldsymbol{A}$, its inverse can be calculated as

$$\boldsymbol{A}^{-1} = \frac{1}{\det(\boldsymbol{A})}\boldsymbol{F}',$$

where $\boldsymbol{F}$ is the matrix of cofactors, i.e., the (i,j) th element of $\boldsymbol{F}$ is the cofactor $(-1)^{i+j}\det(\boldsymbol{A}_{ij})$. The matrix $\boldsymbol{F}'$ is known as the *adjoint* of $\boldsymbol{A}$. For example, when $\boldsymbol{A}$ is 2×2,

$$\boldsymbol{A}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}.$$

Matrix inversion and transposition can be interchanged, i.e., $(\boldsymbol{A}')^{-1} = (\boldsymbol{A}^{-1})'$. For two nonsingular matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ of the same size, we have $\boldsymbol{A}\boldsymbol{B}\boldsymbol{B}^{-1}\boldsymbol{A}^{-1} = \boldsymbol{I}$, so that

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

Some special matrices can be easily inverted. For example, for a diagonal matrix $\boldsymbol{A}$, $\boldsymbol{A}^{-1}$ is also diagonal with the diagonal elements a_{ii}^{-1}; for an orthogonal matrix $\boldsymbol{A}$, $\boldsymbol{A}^{-1} = \boldsymbol{A}'$.

A formula for computing the inverse of a partitioned matrix is

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{F}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{F}^{-1} \\ -\boldsymbol{F}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & \boldsymbol{F}^{-1} \end{bmatrix},$$

where $\boldsymbol{F} = \boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}$, or equivalently,

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{G}^{-1} & -\boldsymbol{G}^{-1}\boldsymbol{B}\boldsymbol{D}^{-1} \\ -\boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{G}^{-1} & \boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{G}^{-1}\boldsymbol{B}\boldsymbol{D}^{-1} \end{bmatrix},$$

where $\boldsymbol{G} = \boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}$, provided that the matrix inverses in the expressions above are well defined. In particular, if this matrix is *block diagonal* so that the off-diagonal blocks are zero matrices, we have

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{A}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}^{-1} \end{bmatrix},$$

provided that $\boldsymbol{A}$ and $\boldsymbol{D}$ are invertible.

1.5 Matrix Rank

The vectors $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ are said to be *linearly independent* if the only solution to

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

is the trivial solution: $c_1 = \cdots = c_n = 0$; otherwise, they are *linearly dependent*. When two (three) vectors are linearly dependent, they are on the same line (plane).

The *column* (*row*) *rank* of a matrix $\boldsymbol{A}$ is the maximum number of linearly independent column (row) vectors of $\boldsymbol{A}$. When the column (row) rank equals the number of column (row) vectors, this matrix is said to be of full column (row) rank. The space *spanned* by the vectors $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ is the collection of all linear combinations of these vectors, denoted as $\text{span}(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)$. The space spanned by the column vectors of $\boldsymbol{A}$ is therefore $\text{span}(\boldsymbol{A})$, which is also known as the *column space* of $\boldsymbol{A}$. A vector $\boldsymbol{z}$ is in $\text{span}(\boldsymbol{A})$ if it can be expressed as $\boldsymbol{Ac}$ for some vector $\boldsymbol{c} \neq \boldsymbol{0}$. Similarly, the space spanned by the row vectors of $\boldsymbol{A}$ is $\text{span}(\boldsymbol{A}')$ and known as the *row space* of $\boldsymbol{A}$. The column (row) rank of $\boldsymbol{A}$ is the *dimension* of the column (row) space of $\boldsymbol{A}$.

Given an $n \times k$ matrix $\boldsymbol{A}$ with $k \leq n$, suppose that $\boldsymbol{A}$ has row rank $r \leq n$ and column rank $c \leq k$. Without loss of generality, assume that the first r row vectors are linear independent. Hence, each row vector $\boldsymbol{a}_i$ can be expressed as

$$\boldsymbol{a}_i = q_{i1}\boldsymbol{a}_1 + q_{i2}\boldsymbol{a}_2 + \cdots + q_{ir}\boldsymbol{a}_r, \qquad i = 1, \ldots, n,$$

with the j^{th} element

$$a_{ij} = q_{i1}a_{1j} + q_{i2}a_{2j} + \cdots + q_{ir}a_{rj}, \qquad i = 1, \ldots, n, \;\; j = 1, \ldots, k.$$

Fixing j, we immediately see that every column vector of $\boldsymbol{A}$ can be written as a linear combination of the vectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_r$. As such, the column rank of $\boldsymbol{A}$ must be less than or equal to r. Similarly, the column rank of $\boldsymbol{A}'$, which is also the row rank of $\boldsymbol{A}$, must be less than or equal to c. This proves the following result.

Lemma 1.3 *The column rank and row rank of a matrix are equal.*

By Lemma 1.3, we can then define the *rank* of $\boldsymbol{A}$ as the maximum number of linearly independent column (or row) vectors of $\boldsymbol{A}$, denoted as $\text{rank}(\boldsymbol{A})$. Clearly, $\text{rank}(\boldsymbol{A}) = \text{rank}(\boldsymbol{A}')$. An $n \times n$ matrix $\boldsymbol{A}$ is said to be of *full rank* if $\text{rank}(\boldsymbol{A}) = n$.

For an $n \times k$ matrix $\boldsymbol{A}$, its *left inverse* is a $k \times n$ matrix $\boldsymbol{A}_L^{-1}$ such that $\boldsymbol{A}_L^{-1}\boldsymbol{A} = \boldsymbol{I}_k$. Similarly, a *right inverse* of $\boldsymbol{A}$ is a $k \times n$ matrix $\boldsymbol{A}_R^{-1}$ such that $\boldsymbol{A}\boldsymbol{A}_R^{-1} = \boldsymbol{I}_n$. The left and right inverses are not unique, however. It can be shown that a matrix possesses a left (right) inverse if and only if it has full column (row) rank. Thus, for a square matrix with full rank, it has both inverses, which are just the unique matrix inverse. Thus, a nonsingular (invertible) matrix must be of full rank and vice versa.

It can be shown that for two $n \times k$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$,

$$\text{rank}(\boldsymbol{A} + \boldsymbol{B}) \leq \text{rank}(\boldsymbol{A}) + \text{rank}(\boldsymbol{B}).$$

If $\boldsymbol{A}$ is $n \times k$ and $\boldsymbol{B}$ is $k \times m$,

$$\operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{B}) - k \leq \operatorname{rank}(\boldsymbol{AB}) \leq \min[\operatorname{rank}(\boldsymbol{A}), \operatorname{rank}(\boldsymbol{B})].$$

For the Kronecker product, we have

$$\operatorname{rank}(\boldsymbol{A} \otimes \boldsymbol{B}) = \operatorname{rank}(\boldsymbol{A}) \operatorname{rank}(\boldsymbol{B}).$$

If $\boldsymbol{A}$ is a nonsingular matrix, we have from the inequality above that

$$\operatorname{rank}(\boldsymbol{AB}) \leq \operatorname{rank}(\boldsymbol{B}) = \operatorname{rank}(\boldsymbol{A}^{-1}\boldsymbol{AB}) \leq \operatorname{rank}(\boldsymbol{AB});$$

i.e., $\operatorname{rank}(\boldsymbol{AB}) = \operatorname{rank}(\boldsymbol{B})$. This also implies that for a nonsingular matrix $\boldsymbol{C}$,

$$\operatorname{rank}(\boldsymbol{BC}) = \operatorname{rank}(\boldsymbol{C}'\boldsymbol{B}') = \operatorname{rank}(\boldsymbol{B}') = \operatorname{rank}(\boldsymbol{B}).$$

Thus, the rank of a matrix is preserved under nonsingular transformations.

Lemma 1.4 *Let $\boldsymbol{A}$ ($n \times n$) and $\boldsymbol{C}$ ($k \times k$) be nonsingular matrices. Then for any $n \times k$ matrix $\boldsymbol{B}$,*

$$\operatorname{rank}(\boldsymbol{B}) = \operatorname{rank}(\boldsymbol{AB}) = \operatorname{rank}(\boldsymbol{BC}).$$

1.6 Eigenvalue and Eigenvector

Given a square matrix $\boldsymbol{A}$, if $\boldsymbol{Ac} = \lambda \boldsymbol{c}$ for some scalar λ and non-zero vector $\boldsymbol{c}$, then $\boldsymbol{c}$ is an *eigenvector* of $\boldsymbol{A}$ corresponding to the *eigenvalue* λ. The system $(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{c} = \boldsymbol{0}$ has a non-trivial solution if and only if

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

This is known as the *characteristic equation* of $\boldsymbol{A}$, from which we can solve for the eigenvalues of $\boldsymbol{A}$. Hence, eigenvalues (eigenvectors) are also referred to as *characteristic roots* (*characteristic vectors*). Note that the eigenvalues and eigenvectors of a real-valued matrix need not be real-valued.

When $\boldsymbol{A}$ is $n \times n$, the characteristic equation is an n^{th}-order polynomial in λ and has at most n distinct solutions. These solutions (eigenvalues) are usually complex-valued. If some eigenvalues take the same value, there may exist several eigenvectors corresponding to the same eigenvalue. Given an eigenvalue λ, let $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_k$ be associated eigenvectors. Then,

$$\boldsymbol{A}(a_1\boldsymbol{c}_1 + a_2\boldsymbol{c}_2 + \ldots + a_k\boldsymbol{c}_k) = \lambda(a_1\boldsymbol{c}_1 + a_2\boldsymbol{c}_2 + \ldots + a_k\boldsymbol{c}_k),$$

so that any linear combination of these eigenvectors is again an eigenvector corresponding to λ. That is, these eigenvectors are closed under scalar multiplication and vector addition and form the *eigenspace* corresponding to λ. As such, for a common eigenvalue, we are mainly concerned with those eigenvectors that are linearly independent.

If $\boldsymbol{A}$ $(n \times n)$ possesses n distinct eigenvalues, each eigenvalue must correspond to one eigenvector, unique up to scalar multiplications. It is therefore typical to normalize eigenvectors such that they have Euclidean norm one. It can also be shown that if the eigenvalues of a matrix are all distinct, their associated eigenvectors must be linearly independent. Let $\boldsymbol{C}$ denote the matrix of these eigenvectors and $\boldsymbol{\Lambda}$ denote the diagonal matrix with diagonal elements being the eigenvalues of $\boldsymbol{A}$. We can write $\boldsymbol{AC} = \boldsymbol{C\Lambda}$. As $\boldsymbol{C}$ is nonsingular, we have

$$\boldsymbol{C}^{-1}\boldsymbol{AC} = \boldsymbol{\Lambda}, \qquad \text{or} \qquad \boldsymbol{A} = \boldsymbol{C\Lambda C}^{-1}.$$

In this case, $\boldsymbol{A}$ is said to be *similar* to $\boldsymbol{\Lambda}$.

When $\boldsymbol{A}$ has n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$, it is readily seen that

$$\det(\boldsymbol{A}) = \det(\boldsymbol{C\Lambda C}^{-1}) = \det(\boldsymbol{\Lambda})\det(\boldsymbol{C})\det(\boldsymbol{C}^{-1}) = \det(\boldsymbol{\Lambda}),$$

and

$$\operatorname{trace}(\boldsymbol{A}) = \operatorname{trace}(\boldsymbol{C\Lambda C}^{-1}) = \operatorname{trace}(\boldsymbol{C}^{-1}\boldsymbol{C\Lambda}) = \operatorname{trace}(\boldsymbol{\Lambda}).$$

This yields the following result.

Lemma 1.5 *Let $\boldsymbol{A}$ be an $n \times n$ matrix with distinct eigenvalues $\lambda_1, \ldots, \lambda_n$. Then*

$$\det(\boldsymbol{A}) = \det(\boldsymbol{\Lambda}) = \prod_{i=1}^{n} \lambda_i,$$
$$\operatorname{trace}(\boldsymbol{A}) = \operatorname{trace}(\boldsymbol{\Lambda}) = \sum_{i=1}^{n} \lambda_i.$$

When $\boldsymbol{A} = \boldsymbol{C\Lambda C}^{-1}$, we have $\boldsymbol{A}^{-1} = \boldsymbol{C\Lambda}^{-1}\boldsymbol{C}^{-1}$. This shows that the eigenvectors of $\boldsymbol{A}^{-1}$ are the same as those of $\boldsymbol{A}$, and the corresponding eigenvalues are simply the reciprocals of the eigenvalues of $\boldsymbol{A}$. Similarly,

$$\boldsymbol{A}^2 = (\boldsymbol{C\Lambda C}^{-1})(\boldsymbol{C\Lambda C}^{-1}) = \boldsymbol{C\Lambda}^2\boldsymbol{C}^{-1},$$

so that the eigenvectors of $\boldsymbol{A}^2$ are the same as those of $\boldsymbol{A}$, and the corresponding eigenvalues are the squares of the eigenvalues of $\boldsymbol{A}$. This result generalizes immediately to $\boldsymbol{A}^k$.

1.7 Symmetric Matrix

More can be said about symmetric matrices. Let $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ be two eigenvectors of $\boldsymbol{A}$ corresponding to the distinct eigenvalues λ_1 and λ_2, respectively. If $\boldsymbol{A}$ is symmetric, then

$$\boldsymbol{c}_2'\boldsymbol{A}\boldsymbol{c}_1 = \lambda_1\boldsymbol{c}_2'\boldsymbol{c}_1 = \lambda_2\boldsymbol{c}_2'\boldsymbol{c}_1.$$

As $\lambda_1 \neq \lambda_2$, it must be true that $\boldsymbol{c}_2'\boldsymbol{c}_1 = 0$, so that they are orthogonal. Given linearly independent eigenvectors that correspond to a common eigenvalue, they can also be orthogonalized. Thus, a symmetric matrix is *orthogonally diagonalizable*, in the sense that

$$\boldsymbol{C}'\boldsymbol{A}\boldsymbol{C} = \boldsymbol{\Lambda}, \qquad \text{or} \qquad \boldsymbol{A} = \boldsymbol{C}\boldsymbol{\Lambda}\boldsymbol{C}',$$

where $\boldsymbol{\Lambda}$ is again the diagonal matrix of the eigenvalues of $\boldsymbol{A}$, and $\boldsymbol{C}$ is the orthogonal matrix of associated eigenvectors.

As nonsingular transformations preserve rank (Lemma 1.4), so do orthogonal transformations. We thus have the result below.

Lemma 1.6 *For a symmetric matrix* $\boldsymbol{A}$*,* $\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{\Lambda})$*, the number of non-zero eigenvalues of* $\boldsymbol{A}$*.*

Moreover, when $\boldsymbol{A}$ is diagonalizable, the assertions of Lemma 1.5 remain valid, whether or not the eigenvalues of $\boldsymbol{A}$ are distinct.

Lemma 1.7 *Let* $\boldsymbol{A}$ *be an* $n \times n$ *symmetric matrix. Then,*

$$\det(\boldsymbol{A}) = \det(\boldsymbol{\Lambda}) = \prod_{i=1}^{n} \lambda_i,$$

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

By Lemma 1.7, a symmetric matrix is nonsingular if its eigenvalues are all non-zero.

A symmetric matrix $\boldsymbol{A}$ is said to be *positive definite* if $\boldsymbol{b}'\boldsymbol{A}\boldsymbol{b} > 0$ for all vectors $\boldsymbol{b} \neq \boldsymbol{0}$; $\boldsymbol{A}$ is said to be *positive semi-definite* if $\boldsymbol{b}'\boldsymbol{A}\boldsymbol{b} \geq 0$ for all $\boldsymbol{b} \neq \boldsymbol{0}$. A positive definite matrix thus must be nonsingular, but a positive semi-definite matrix may be singular. Suppose that $\boldsymbol{A}$ is a symmetric matrix orthogonally diagonalized as $\boldsymbol{C}'\boldsymbol{A}\boldsymbol{C} = \boldsymbol{\Lambda}$. If $\boldsymbol{A}$ is also positive semi-definite, then for any $\boldsymbol{b} \neq \boldsymbol{0}$,

$$\boldsymbol{b}'\boldsymbol{\Lambda}\boldsymbol{b} = \boldsymbol{b}'(\boldsymbol{C}'\boldsymbol{A}\boldsymbol{C})\boldsymbol{b} = \tilde{\boldsymbol{b}}'\boldsymbol{A}\tilde{\boldsymbol{b}} \geq 0,$$

where $\tilde{\boldsymbol{b}} = \boldsymbol{C}\boldsymbol{b}$. This shows that $\boldsymbol{\Lambda}$ is also positive semi-definite, and all the diagonal elements of $\boldsymbol{\Lambda}$ must be non-negative. It can be seen that the converse also holds.

Lemma 1.8 *A symmetric matrix is positive definite (positive semi-definite) if, and only if, its eigenvalues are all positive (non-negative).*

For a symmetric and positive definite matrix $\boldsymbol{A}$, $\boldsymbol{A}^{-1/2}$ is such that $\boldsymbol{A}^{-1/2\prime}\boldsymbol{A}^{-1/2} = \boldsymbol{A}^{-1}$. In particular, by orthogonal diagonalization,

$$\boldsymbol{A}^{-1} = \boldsymbol{C}\boldsymbol{\Lambda}^{-1}\boldsymbol{C}' = (\boldsymbol{C}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{C}')(\boldsymbol{C}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{C}'),$$

so that we may choose $\boldsymbol{A}^{-1/2} = \boldsymbol{C}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{C}'$. The inverse of $\boldsymbol{A}^{-1/2}$ is $\boldsymbol{A}^{1/2} = \boldsymbol{C}\boldsymbol{\Lambda}^{1/2}\boldsymbol{C}'$. It follows that $\boldsymbol{A}^{1/2}\boldsymbol{A}^{1/2\prime} = \boldsymbol{A}$, and $\boldsymbol{A}^{-1/2}\boldsymbol{A}\boldsymbol{A}^{-1/2\prime} = \boldsymbol{I}$. Note that $\boldsymbol{\Lambda}^{-1/2}\boldsymbol{C}'$ is also a legitimate choice of $\boldsymbol{A}^{-1/2}$, yet it is not symmetric.

Finally, we know that for two positive real numbers a and b, $a \geq b$ implies $b^{-1} \geq a^{-1}$. This result can be generalized to compare two positive definite matrices, as stated below without proof.

Lemma 1.9 *Given two symmetric and positive definite matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, if $\boldsymbol{A} - \boldsymbol{B}$ is positive semi-definite, then so is $\boldsymbol{B}^{-1} - \boldsymbol{A}^{-1}$.*

1.8 Orthogonal Projection

A matrix $\boldsymbol{A}$ is said to be *idempotent* if $\boldsymbol{A}^2 = \boldsymbol{A}$. Given a vector $\boldsymbol{y}$ in the Euclidean space V, a *projection* of $\boldsymbol{y}$ onto a subspace S of V is a linear transformation of $\boldsymbol{y}$ to S. The resulting projected vector can be written as $\boldsymbol{P}\boldsymbol{y}$, where $\boldsymbol{P}$ is the associated transformation matrix. Given the projection $\boldsymbol{P}\boldsymbol{y}$ in S, further projection to S should have no effect on $\boldsymbol{P}\boldsymbol{y}$, i.e.,

$$\boldsymbol{P}(\boldsymbol{P}\boldsymbol{y}) = \boldsymbol{P}^2\boldsymbol{y} = \boldsymbol{P}\boldsymbol{y}.$$

Thus, a matrix $\boldsymbol{P}$ is said to be a *projection matrix* if it is *idempotent.*

A projection of $\boldsymbol{y}$ onto S is orthogonal if the projection $\boldsymbol{P}\boldsymbol{y}$ is orthogonal to the difference between $\boldsymbol{y}$ and $\boldsymbol{P}\boldsymbol{y}$. That is,

$$(\boldsymbol{y} - \boldsymbol{P}\boldsymbol{y})'\boldsymbol{P}\boldsymbol{y} = \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P})'\boldsymbol{P}\boldsymbol{y} = \boldsymbol{0}.$$

As $\boldsymbol{y}$ is arbitrary, the equality above holds if, and only if, $(\boldsymbol{I} - \boldsymbol{P})'\boldsymbol{P} = \boldsymbol{0}$. Consequently, $\boldsymbol{P} = \boldsymbol{P}'\boldsymbol{P}$ and $\boldsymbol{P}' = \boldsymbol{P}'\boldsymbol{P}$. This shows that $\boldsymbol{P}$ must be symmetric. Thus, a matrix is an *orthogonal projection matrix* if, and only if, it is symmetric and idempotent. It can be easily verified that the orthogonal projection $\boldsymbol{P}\boldsymbol{y}$ must be unique.

When $\boldsymbol{P}$ is an orthogonal projection matrix, it is easily seen that $\boldsymbol{I} - \boldsymbol{P}$ is idempotent because

$$(\boldsymbol{I} - \boldsymbol{P})^2 = \boldsymbol{I} - 2\boldsymbol{P} + \boldsymbol{P}^2 = \boldsymbol{I} - \boldsymbol{P}.$$

As $\boldsymbol{I} - \boldsymbol{P}$ is also symmetric, it is an orthogonal projection matrix. Since $(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{P} = \boldsymbol{0}$, the projections $\boldsymbol{P}\boldsymbol{y}$ and $(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{y}$ must be orthogonal. This shows that any vector $\boldsymbol{y}$ can be uniquely decomposed into two orthogonal components:

$$\boldsymbol{y} = \boldsymbol{P}\boldsymbol{y} + (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{y}.$$

Define the *orthogonal complement* of a subspace $S \subseteq V$ as

$$S^{\perp} = \{\boldsymbol{v} \in V : \boldsymbol{v}'\boldsymbol{s} = 0, \text{ for all } \boldsymbol{s} \in S\}.$$

If $\boldsymbol{P}$ is the orthogonal projection matrix that projects vectors onto $S \subseteq V$, we have $\boldsymbol{P}\boldsymbol{s} = \boldsymbol{s}$ for any $\boldsymbol{s} \in S$. It follows that $(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{y}$ is orthogonal to $\boldsymbol{s}$ and that $(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{y}$ is the orthogonal projection of $\boldsymbol{y}$ onto $S^{\perp}$.

Intuitively, the orthogonal projection $\boldsymbol{P}\boldsymbol{y}$ can be interpreted as the "best approximation" of $\boldsymbol{y}$ in S, in the sense that $\boldsymbol{P}\boldsymbol{y}$ is the closest to $\boldsymbol{y}$ in terms of the Euclidean norm. To see this, we observe that for any $\boldsymbol{s} \in S$,

$$\begin{aligned} \|\boldsymbol{y} - \boldsymbol{s}\|^2 &= \|\boldsymbol{y} - \boldsymbol{P}\boldsymbol{y} + \boldsymbol{P}\boldsymbol{y} - \boldsymbol{s}\|^2 \\ &= \|\boldsymbol{y} - \boldsymbol{P}\boldsymbol{y}\|^2 + \|\boldsymbol{P}\boldsymbol{y} - \boldsymbol{s}\|^2 + 2(\boldsymbol{y} - \boldsymbol{P}\boldsymbol{y})'(\boldsymbol{P}\boldsymbol{y} - \boldsymbol{s}) \\ &= \|\boldsymbol{y} - \boldsymbol{P}\boldsymbol{y}\|^2 + \|\boldsymbol{P}\boldsymbol{y} - \boldsymbol{s}\|^2. \end{aligned}$$

This establishes the following result.

Lemma 1.10 *Let $\boldsymbol{y}$ be a vector in V and $\boldsymbol{P}\boldsymbol{y}$ its orthogonal projection onto $S \subseteq V$. Then,*

$$\|\boldsymbol{y} - \boldsymbol{P}\boldsymbol{y}\| \leq \|\boldsymbol{y} - \boldsymbol{s}\|,$$

for all $\boldsymbol{s} \in S$.

Let $\boldsymbol{A}$ be a symmetric and idempotent matrix and $\boldsymbol{C}$ be the orthogonal matrix that diagonalizes $\boldsymbol{A}$ to $\boldsymbol{\Lambda}$. Then,

$$\boldsymbol{\Lambda} = \boldsymbol{C}'\boldsymbol{A}\boldsymbol{C} = \boldsymbol{C}'\boldsymbol{A}(\boldsymbol{C}\boldsymbol{C}')\boldsymbol{A}\boldsymbol{C} = \boldsymbol{\Lambda}^2.$$

This is possible only when the eigenvalues of $\boldsymbol{A}$ are zero and one. The result below now follows from Lemmas 1.8.

Lemma 1.11 *A symmetric and idempotent matrix is positive semi-definite with the eigenvalues 0 and 1.*

Moreover, trace($\boldsymbol{\Lambda}$) is the number of non-zero eigenvalues of $\boldsymbol{A}$ and hence rank($\boldsymbol{\Lambda}$). When $\boldsymbol{A}$ is symmetric, rank($\boldsymbol{A}$) = rank($\boldsymbol{\Lambda}$) by Lemma 1.6, and trace($\boldsymbol{A}$) = trace($\boldsymbol{\Lambda}$) by Lemma 1.7. Combining these results we have:

Lemma 1.12 *For a symmetric and idempotent matrix* $\boldsymbol{A}$, $\text{rank}(\boldsymbol{A}) = \text{trace}(\boldsymbol{A})$, *the number of non-zero eigenvalues of* $\boldsymbol{A}$.

Given an $n \times k$ matrix $\boldsymbol{A}$, it is easy to see that $\boldsymbol{A}'\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}'$ are symmetric and positive semi-definite. Let $\boldsymbol{x}$ denote a vector orthogonal to the rows of $\boldsymbol{A}'\boldsymbol{A}$; i.e., $\boldsymbol{A}'\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$. Hence $\boldsymbol{x}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{x} = 0$, so that $\boldsymbol{A}\boldsymbol{x}$ must be a zero vector. That is, $\boldsymbol{x}$ is also orthogonal to the rows of $\boldsymbol{A}$. Conversely, $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ implies $\boldsymbol{A}'\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$. This shows that the orthogonal complement of the row space of $\boldsymbol{A}$ is the same as the orthogonal complement of the row space of $\boldsymbol{A}'\boldsymbol{A}$. Hence, these two row spaces are also the same. Similarly, the column space of $\boldsymbol{A}$ is the same as the column space of $\boldsymbol{A}\boldsymbol{A}'$. It follows from Lemma 1.3 that

$$\text{rank}(\boldsymbol{A}) = \text{rank}(\boldsymbol{A}'\boldsymbol{A}) = \text{rank}(\boldsymbol{A}\boldsymbol{A}').$$

In particular, if $\boldsymbol{A}$ $(n \times k)$ is of full column rank $k < n$, then $\boldsymbol{A}'\boldsymbol{A}$ is $k \times k$ and hence of full rank k (nonsingular), but $\boldsymbol{A}\boldsymbol{A}'$ is $n \times n$ and hence singular. The result below is now immediate.

Lemma 1.13 *If* $\boldsymbol{A}$ *is an* $n \times k$ *matrix with full column rank* $k < n$, *then,* $\boldsymbol{A}'\boldsymbol{A}$ *is symmetric and positive definite.*

Given an $n \times k$ matrix $\boldsymbol{A}$ with full column rank $k < n$, $\boldsymbol{P} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$ is clearly symmetric and idempotent and hence an orthogonal projection matrix. As

$$\text{trace}(\boldsymbol{P}) = \text{trace}(\boldsymbol{A}'\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}) = \text{trace}(\boldsymbol{I}_k) = k,$$

we have from Lemmas 1.11 and 1.12 that $\boldsymbol{P}$ has exactly k eigenvalues equal to 1 and that $\text{rank}(\boldsymbol{P}) = k$. Similarly, $\text{rank}(\boldsymbol{I} - \boldsymbol{P}) = n - k$. Moreover, any vector $\boldsymbol{y} \in \text{span}(\boldsymbol{A})$ can be written as $\boldsymbol{A}\boldsymbol{b}$ for some non-zero vector $\boldsymbol{b}$, and

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

This suggests that $\boldsymbol{P}$ must project vectors onto $\text{span}(\boldsymbol{A})$. On the other hand, when $\boldsymbol{y} \in \text{span}(\boldsymbol{A})^{\perp}$, $\boldsymbol{y}$ is orthogonal to the column vectors of $\boldsymbol{A}$ so that $\boldsymbol{A}'\boldsymbol{y} = \boldsymbol{0}$. It follows that $\boldsymbol{P}\boldsymbol{y} = \boldsymbol{0}$ and $(\boldsymbol{I} - \boldsymbol{P})\boldsymbol{y} = \boldsymbol{y}$. Thus, $\boldsymbol{I} - \boldsymbol{P}$ must project vectors onto $\text{span}(\boldsymbol{A})^{\perp}$. These results are summarized below.

Lemma 1.14 *Let* $\boldsymbol{A}$ *be an* $n \times k$ *matrix with full column rank* k. *Then,* $\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$ *orthogonally projects vectors onto* $\text{span}(\boldsymbol{A})$ *and has rank* k; $\boldsymbol{I}_n - \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$ *orthogonally projects vectors onto* $\text{span}(\boldsymbol{A})^{\perp}$ *and has rank* $n - k$.

References

Anton, Howard (1981). *Elementary Linear Algebra*, third edition, New York: Wiley.

Basilevsky, Alexander (1983). *Applied Matrix Algebra in the Statistical Sciences*, New York: North-Holland.

Graybill, Franklin A. (1969). *Introduction to Matrices with Applications in Statistics*, second edition, Belmont, CA: Wadsworth.

Noble, Ben and James W. Daniel. *Applied Linear Algebra*, second edition, Englewood Cliffs, NJ: Prentice-Hall.