1 Review of Matrix Algebra

There are numerous excellent books on matrix algebra that contain results useful for the presentation in this book. See for instance the books by Graybill (1961), Horn and Johnson (1985), Magnus and Neudecker (1999), Abadir and Magnus (2005), and Bernstein (2005).

1.1 Introduction

Definition 1.1 *An $m \times n$ matrix A is a rectangular array of elements in m rows and n columns.*

Let a_{ij} be the ith row and jth column of A. Then A can be represented as

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} = (a_{ij}).$$

To indicate an $m \times n$ matrix A, we sometimes write $A : m \times n$ or $\underset{m\times n}{A}$. A matrix with $m = n$ rows and columns is called a *square matrix*. A square matrix having zeros as elements below (above) the diagonal is called an upper (lower) *triangular matrix*.

For any scalar c, $cA = (ca_{ij})$. If $B = (b_{ij})$ is also an $m \times n$ matrix, then $A + B = (a_{ij} + b_{ij})$. If $C = (c_{ij})$ is an $n \times p$ matrix, then $AC = \left(\sum_{k=1}^{n} a_{ik} c_{kj}\right)$. For three matrices A, B, and C, it is easy to verify that

$$(AB)\,C = A\,(BC) = ABC,$$

as long as the above calculation makes sense. Nevertheless, $AB \neq BA$ in general.

Definition 1.2 *The transpose $A^T : n \times m$ of a matrix $A : m \times n$ is obtained by interchanging the rows and columns of A. Thus*

$$A^T = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix}.$$

If $A = (a_{ij})$, frequently we write $A^T = (a_{ji})$. It is easy to verify that the transpose satisfies the following properties:

$$\left(A^T\right)^T = A, \;\; (A+B)^T = A^T + B^T, \;\; (AB)^T = B^T A^T.$$

A square matrix is called *symmetric* if $A^T = A$.

Definition 1.3 *An $m \times 1$ matrix a is said to be a column m-vector and written as*

$$a = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}.$$

A $1 \times m$ matrix a^T is said to be a row vector and written as $a^T = (a_1, \cdots, a_m)$.

Unless otherwise specified, all vectors in this appendix are column vectors. The n-vector

$$e_i = (0, \cdots, 0, 1, 0, \cdots, 0)^T$$

with the ith element as 1 and others as 0, is called the ith *unit vector*. An $n \times n$ square matrix with 1 on the diagonal and 0 elsewhere is called an *identity matrix*, and frequently denoted as I_n. Two $m \times 1$ vectors a and b are *orthogonal* if $a^T b = 0$. An $n \times n$ square matrix is orthogonal if $A^T A = I_n$.

Definition 1.4 *A matrix A is said to be partitioned if its elements are arranged in submatrices.*

Let

$$\underset{m\times n}{A} = \begin{pmatrix} \underset{m_1\times n_1}{A_{11}} & \underset{m_1\times n_2}{A_{12}} \\ \underset{m_2\times n_1}{A_{21}} & \underset{m_2\times n_2}{A_{22}} \end{pmatrix} \text{ and } \underset{m\times n}{B} = \begin{pmatrix} \underset{m_1\times n_1}{B_{11}} & \underset{m_1\times n_2}{B_{12}} \\ \underset{m_2\times n_1}{B_{21}} & \underset{m_2\times n_2}{B_{22}} \end{pmatrix},$$

where $m = m_1 + m_2$, and $n = n_1 + n_2$. Then

(i) $A + B = \begin{pmatrix} A_{11} + B_{11} & A_{12} + B_{12} \\ A_{21} + B_{21} & A_{22} + B_{22} \end{pmatrix}$.

(ii) $A^T = \begin{pmatrix} A_{11}^T & A_{21}^T \\ A_{12}^T & A_{22}^T \end{pmatrix}$.

1.2 Trace of a Matrix

Definition 1.5 *For an $n \times n$ square matrix $A = (a_{ij})$, its trace is defined as the sum of its diagonal elements, i.e.,*

$$tr(A) = \sum_{i=1}^{n} a_{ii}.$$

Let A and B be both $n \times n$ square matrices, and let c be a scalar. Then we have the following rules:

(i) $\mathrm{tr}(A \pm B) = \mathrm{tr}(A) \pm \mathrm{tr}(B)$;

(ii) $\mathrm{tr}(A^T) = \mathrm{tr}(A)$;

(iii) $\mathrm{tr}(cA) = c\mathrm{tr}(A)$;

(iv) $\mathrm{tr}(AB) = \mathrm{tr}(BA)$;

(v) $\mathrm{tr}(AA^T) = \mathrm{tr}(A^T A) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2$.

Note that rules (iv) and (v) also hold for the case $A : m \times n$ and $B : n \times m$.

1.3 Determinant of a Matrix

Definition 1.6 *Let $n > 1$. The determinant of an $n \times n$ square matrix A is defined by*

$$|A| = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} |M_{ij}| \text{ for any fixed } j,$$

with $|M_{ij}|$ being the minor of the elements a_{ij}, i.e., $|M_{ij}|$ is the determinant of the remaining $(n-1) \times (n-1)$ matrix when the ith row and the jth column of A are deleted. $A_{ij} = (-1)^{i+j} |M_{ij}|$ is called the cofactor of a_{ij}.

Example 1.7 *If $A = (a_{ij})$ is a 2×2 matrix, $|A| = a_{11}a_{22} - a_{12}a_{21}$. If $A = (a_{ij})$ is a 3×3 matrix, we can hold column $j = 1$ fixed and have*

$$\begin{aligned}
A_{11} &= (-1)^{1+1} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} = M_{11}, \\
A_{21} &= (-1)^{2+1} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} = -M_{21}, \\
A_{31} &= (-1)^{3+1} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} = M_{31}.
\end{aligned}$$

Consequently, $|A| = a_{11}A_{11} + a_{21}A_{21} + a_{31}A_{31}$.

A square matrix A is said to be *nonsingular* if $|A| \neq 0$. Otherwise, A is said to be *singular*.

Let A and B be $n \times n$ square matrices, and c be a scalar. Then we have the following properties.

(*i*) $\left|A^T\right| = |A|$;

(*ii*) $|cA| = c^n |A|$;

(*iii*) $|AB| = |A| |B|$;

(*iv*) $\left|A^2\right| = |A|^2$;

(*v*) If $A = (a_{ij})$ is diagonal or triangular, then $|A| = \Pi_{i=1}^{n} a_{ii}$;

(*vi*) If $A = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix}$, then $|A| = |A_{11}| |A_{22}|$, where A_{11} is also a square matrix.

1.4 Rank of a Matrix

Definition 1.8 *A set of vectors $v_1, ..., v_n$ are linearly dependent if only if any one of the vectors in the set can be expressed as a linear combination of the others. They are linearly independent if and only if the only solution to $c_1v_1 + \cdots + c_nv_n = 0$ is $c_1 = \cdots = c_n = 0$.*

Definition 1.9 *The rank of $A : m \times n$ is the maximum number of linearly independent rows (or columns) of A, denoted as rank(A).*

Let A be an $m \times n$ matrix. Then

(*i*) $0 \leq \text{rank}(A) \leq \min\{m, n\}$;

(*ii*) $\text{rank}(A) = \text{rank}(A^T)$;

(*iii*) $\text{rank}(A + B) \leq \text{rank}(A) + \text{rank}(B)$;

(*iv*) $\text{rank}(AB) \leq \min\{\text{rank}(A), \text{rank}(B)\}$;

(*v*) $\text{rank}(A) = \text{rank}(A^TA) = \text{rank}(AA^T)$;

(*vi*) For nonsingular $B : m \times m$ and $C : n \times n$, we have $\text{rank}(BAC) = \text{rank}(A)$;

(*vii*) If $A : n \times n$ is diagonal, then $\text{rank}(A)$ equals the number of a_{ii} that is nonzero.

1.5 Inverse of a Matrix

Definition 1.10 *Let A be an $n \times n$ square matrix with full rank, the inverse A^{-1} of A is defined to be a matrix B satisfying*

$$AB = BA = I_n.$$

The inverse A^{-1} exists if and only if A is full rank, or equivalently, A is nonsingular.

Example 1.11 *If $A = (a_{ij})$ is a 2×2 nonsingular matrix, then*

$$A^{-1} = \frac{1}{|A|} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix},$$

where recall $|A| = a_{11}a_{22} - a_{12}a_{21}$.

Let A and B be $n \times n$ square matrices with full rank, and c be a nonzero scalar. Then we have the following properties.

(i) $(cA)^{-1} = c^{-1}A^{-1}$;
(ii) $(AB)^{-1} = B^{-1}A^{-1}$;
(iii) $\left(A^{-1}\right)^{-1} = A$;
(iv) $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$;
(v) $\left|A^{-1}\right| = |A|^{-1}$;
(vi) If $A = \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix}$, then $A^{-1} = \begin{pmatrix} A_{11}^{-1} & 0 \\ 0 & A_{22}^{-1} \end{pmatrix}$ where A_{11} is also a square matrix.

Frequently we also use the generalized inverse of a matrix. A *generalized inverse* A^- of a matrix A satisfies the property

$$A^- A A^- = A^-.$$

Note that A^- is generally not unique and it reduces to the usual inverse A^{-1} if A is a nonsingular square matrix. The *Moore-Penrose generalized inverse* A^- exists, is unique, and satisfies the following three properties:

$$(i)\ A^- A A^- = A^-,\ \ (ii)\ \ AA^- \text{is symmetric, and}\ \ (iii)\ \ A^- A \text{ is symmetric.}$$

1.6 Eigenvalues and Eigenvectors

Definition 1.12 *Let A be an $n \times n$ square matrix, then*

$$q(\lambda) = |A - \lambda I_n|$$

is an nth order polynomial in λ. The n roots $\lambda_1, ..., \lambda_n$ of the characteristic function $q(\lambda) = |A - \lambda I_n| = 0$ are called eigenvalues or characteristic roots of A.

Let A be an $n \times n$ square matrix and x be a n-vector. Consider

$$Ax = \lambda x.$$

We have

$$(A - \lambda I_n)\, x = 0.$$

The nontrivial solution of x (i.e., $x \neq 0$) to the above problem exists only if

$$|A - \lambda I_n| = 0.$$

(If $|A - \lambda I_n| \neq 0$, then $(A - \lambda I_n)^{-1}$ exists so that $x = (A - \lambda I_n)^{-1} 0 = 0$.)

Definition 1.13 *Let λ^* be an eigenvalue of A. Corresponding to λ^*, the value of x^* that satisfies $Ax^* = \lambda^* x^*$ is called the eigenvector of A. Often, we impose the normalization rule ${x^*}^T x^* = 1$.*

Let A be a real symmetric $n \times n$ matrix.

(i) The eigenvalues of A are real.

(ii) The eigenvectors corresponding to distinct eigenvalues are pairwise orthogonal.

(iii) A can be diagonalized. That is, there exists an *orthogonal matrix* X (i.e., $X^T X = XX^T = I_n$, or equivalently, $X^T = X^{-1}$) and a diagonal matrix

$$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

such that

$$X^T A X = \Lambda.$$

(iv) $|A| = \Pi_{i=1}^{n} \lambda_i$.

(v) $\mathrm{tr}(A) = \sum_{i=1}^{n} \lambda_i$.

1.7 Definite Matrices and Quadratic Forms

Definition 1.14 *Let A be an $n \times n$ symmetric matrix and x an $n \times 1$ vector. Then the quadratic form in x is defined as the function*

$$Q(x) = x^T A x = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j.$$

A is positive definite (p.d.) if $x^T Ax > 0$ for all $x \neq 0$. It is negative definite (n.d.) if $x^T Ax < 0$ for all $x \neq 0$. It is positive semidefinite. (p.s.d.) if $x^T Ax \geq 0$ for all x. It is negative semidefinite. (n.s.d.) if $x^T Ax \leq 0$ for all x.

We now state some properties about definite matrices.

(i) Let a be an $n \times 1$ vector, then $A = aa^T$ is always p.s.d.

(ii) If A is p.s.d. (p.d.), then the eigenvalues of A are all not less than 0 (greater than 0).

(iii) Let A be a real symmetric p.s.d. $n \times n$ matrix. Then there exists a matrix C such that

$$A = C^T C.$$

Note that C is not unique in (iii). Since A is real symmetric, there exists an *orthogonal matrix* X and a nonnegative diagonal matrix Λ such that

$$A = X\Lambda X^T = X\Lambda^{1/2}\Lambda^{1/2}X^T = C^T C,$$

where

$$\Lambda^{1/2} = \begin{bmatrix} \sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{\lambda_n} \end{bmatrix}$$

and $C = \Lambda^{1/2}X^T$. Another choice of C is $C = X\Lambda^{1/2}X^T$. To see so, note that

$$C^T C = X\Lambda^{1/2}X^T X\Lambda^{1/2}X^T = X\Lambda^{1/2}\Lambda^{1/2}X^T = X\Lambda X^T = A.$$

If we require C to be real symmetric, then C is unique.

1.8 Idempotent Matrix

Definition 1.15 *An $n \times n$ square matrix A is idempotent if and only if $A^2 \equiv AA = A$. An idempotent matrix A is called an orthogonal projector or a projection matrix if $A = A^T$.*

Let A be an $n \times n$ idempotent matrix. Then we have:

(i) $\mathrm{rank}(A) = \mathrm{tr}(A)$.

(ii) $I_n - A$ is idempotent too.

(iii) If A is symmetric, then it is p.s.d.

(iv) If A is symmetric, then its eigenvalues are 0 or 1.

(v) If A is of full rank n, then $A = I_n$.

(vi) If A and B are idempotent and if $AB = BA$, then AB is also idempotent.

(vii) If A is idempotent and B is orthogonal, then BAB^T is also idempotent.

To see (iv), we start with the relationship: $Ax = \lambda x$. The idempotence of A implies that

$$Ax = AAx = \lambda Ax = \lambda^2 x.$$

It follows that

$$x^T Ax = \lambda x^T x = \lambda^2 x^T x,$$

and thus $\lambda(1-\lambda)x^T x = 0$. So $\lambda = 0$ or $\lambda = 1$.

Note that a matrix can be idempotent but not symmetric, e.g.,

$$A = \begin{pmatrix} -2 & 1 \\ -6 & 3 \end{pmatrix}.$$

1.9 Differentiation of Matrices

Definition 1.16 *If $f(X)$ is a real function of an $m \times n$ matrix $X = (x_{ij})$, then the partial differential of f with respect to X is defined as the $m \times n$ matrix of partial differentials $\partial f / \partial x_{ij}$:*

$$\frac{\partial f(X)}{\partial X} = \begin{pmatrix} \frac{\partial f(X)}{\partial x_{11}} & \cdots & \frac{\partial f(X)}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(X)}{\partial x_{m1}} & \cdots & \frac{\partial f(X)}{\partial x_{mn}} \end{pmatrix}.$$

In particular, let $x = (x_1, ..., x_n)^T$ and $y = f(x)$ be a real function. Then

$$\frac{\partial f(x)}{\partial x} = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix},$$

and

$$\frac{\partial f(x)}{\partial x^T} = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} & \cdots & \frac{\partial f(x)}{\partial x_n} \end{pmatrix}.$$

Definition 1.17 *Let $x = (x_1, ..., x_n)^T$ and $g(x) = (g_1(x), .., g_m(x))^T$. We define*

$$\frac{\partial g(x)}{\partial x^T} = \begin{pmatrix} \frac{\partial g_1(x)}{\partial x^T} \\ \vdots \\ \frac{\partial g_m(x)}{\partial x^T} \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1(x)}{\partial x_1} & \cdots & \frac{\partial g_1(x)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m(x)}{\partial x_1} & \cdots & \frac{\partial g_m(x)}{\partial x_n} \end{pmatrix},$$

and

$$\frac{\partial g^T(x)}{\partial x} = \left(\frac{\partial g(x)}{\partial x^T}\right)^T.$$

We now apply the above definitions to some specific functions of matrices.

Example 1.18 *Let $a = (a_1, ..., a_n)^T$, $x = (x_1, ..., x_n)^T$, and $y = a^T x = \sum_{i=1}^n a_i x_i$. Then*

$$\frac{\partial \left(a^T x\right)}{\partial x} = \begin{pmatrix} \frac{\partial \left(a^T x\right)}{\partial x_1} \\ \vdots \\ \frac{\partial \left(a^T x\right)}{\partial x_n} \end{pmatrix} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = a.$$

Example 1.19 *Further, let $A = (a_{ij})$ be a $m \times n$ matrix and $y = Ax$. Then*

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_1^T \\ \vdots \\ a_m^T \end{pmatrix} x = \begin{pmatrix} a_1^T x \\ \vdots \\ a_m^T x \end{pmatrix},$$

where a_i^T is the ith row of A. Then $\partial y_i / \partial x^T = a_i^T$. So

$$\frac{\partial (Ax)}{\partial x^T} = \begin{pmatrix} \frac{\partial y_1}{\partial x^T} \\ \vdots \\ \frac{\partial y_m}{\partial x^T} \end{pmatrix} = \begin{pmatrix} a_1^T \\ \vdots \\ a_m^T \end{pmatrix} = A.$$

Similarly,

$$\frac{\partial\left(x^{T}A^{T}\right)}{\partial x}=A^{T}.$$

Example 1.20 *Let* $x=(x_1,...,x_n)$, $y=(y_1,...,y_n)$, *and* $A=(a_{ij})$ *be* $n\times n$ *square matrix. Let* $z=x^{T}Ay=\sum_{i=1}^{n}\sum_{j=1}^{n}x_iy_ja_{ij}$. *Then*

$$\frac{\partial\left(x^{T}Ay\right)}{\partial x}=Ay.$$

In particular, if $y=x$, *we have*

$$\begin{aligned}\frac{\partial\left(x^{T}Ax\right)}{\partial x} &= \left(A+A^{T}\right)x\\ &= 2Ax \text{ if } A \text{ is symmetric.}\end{aligned}$$

Noticing that $\partial\left(x^{T}Ax\right)/\partial a_{ij}=x_ix_j$, *we also have*

$$\frac{\partial\left(x^{T}Ax\right)}{\partial A}=xx^{T}.$$

References

[1] Abadir, K. M., and Magnus, J. R., 2005. *Matrix Algebra.* Cambridge University Press, Cambridge.

[2] Bernstein, D. S., 2005. *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear System Theory.* Princeton University Press, Princeton.

[3] Graybill, F. A., 1961. *An Introduction to Linear Statistical Models.* McGraw-Hill, New York.

[4] Horn, R. A. and Johnson, C. R., 1985. *Matrix Analysis.* Cambridge University Press, Cambridge.

[5] Magnus, J. R. and Neudecker, H., 1999. *Matrix Differential Calculus with Applications in Statistics and Econometrics.* John Willey & Sons, New York.