Review of Matrix Algebra

1 Introduction

An matrix $\boldsymbol{A}$ of *dimension* $m \times n$ is a rectangular array of numbers with m rows and n columns, with the element in the ith row and jth being denoted by a_{ij}, $i = 1, \ldots, m$, $j = 1, \ldots, n$, that is

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

We may also write matrix $\boldsymbol{A}$ in (1) in terms of its typical elements a_{ij} as $\boldsymbol{A} = [a_{ij}]_{i=1,\ldots,m;j=1,\ldots,n}$ or simply $[a_{ij}]$ if there is no ambiguity concerning its dimension.

1.1 Basic Definitions

(1) The *transpose* of the $m \times n$ matrix $\boldsymbol{A}$, denoted by $\boldsymbol{A}'$ (A *prime*), is the matrix obtained by interchanging the rows and columns of $\boldsymbol{A}$, that is, if $\boldsymbol{A}$ is as in (1), then $\boldsymbol{A}'$ is the $n \times m$ matrix with typical element a'_{ij} given by

$$\boldsymbol{A}' = [a'_{ij}] = [a_{ji}]. \tag{2}$$

A matrix is equal to the transpose of its own transpose: $(\boldsymbol{A}')' = \boldsymbol{A}$.

(2) A *square matrix* is a matrix where $m = n$.

(3) A square matrix is *symmetric* if $\boldsymbol{A} = \boldsymbol{A}'$, that is, $a_{ij} = a_{ji}$. For example,

$$\boldsymbol{A} = \begin{bmatrix} 3 & 1 & -1 \\ 1 & 3 & -1 \\ -1 & -1 & 5 \end{bmatrix} \tag{3}$$

is a symmetric 3×3 matrix.

(4) A square matrix with all off–diagonal elements being zero is a *diagonal matrix*, i.e., for such a matrix

$$a_{ij} = \begin{cases} a_{ii} & \text{for i = j} \\ 0 & \text{for } i \neq j. \end{cases} \tag{4}$$

A special case of a diagonal matrix is the *identity matrix* of dimension n, $\boldsymbol{I}_n$, where all the diagonal elements are equal to one, i.e.,

$$\boldsymbol{I}_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{5}$$

(5) A matrix with a single row is a *row vector*, and a matrix with a single column is a *column vector*.

1.2 Elementary operations: Summation and Multiplication

(1) Summation: If $\boldsymbol{A}$ and $\boldsymbol{B}$ are of the same order $m \times n$, then $m \times n$ matrix $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$ is defined by

$$\boldsymbol{C} = [c_{ij}] = [a_{ij} + b_{ij}]. \tag{6}$$

(3) Scalar multiplication: For $\alpha \in \mathbb{R}$, $\alpha \boldsymbol{A} = \alpha[a_{ij}] = [\alpha a_{ij}]$.

(2) Matrix multiplication: Consider two n–dimensional vectors $\boldsymbol{a} = [a_1, \ldots, a_n]$ and $\boldsymbol{b} = [b_1, \ldots, b_n]$. The *inner product* of these vectors is given by

$$\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \sum_{i=1}^{n} a_i b_i. \tag{7}$$

Now consider matrices $\boldsymbol{A}_{m\times n}$ and $\boldsymbol{B}_{n\times p}$ of dimensions $m \times n$ and $n \times p$, respectively, where it is crucial that the number of columns of $\boldsymbol{A}$ is equal to the number of rows of $\boldsymbol{B}$. Write these matrices in terms of their row and column vectors, respectively, as

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_m \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \dots & \boldsymbol{b}_p \end{bmatrix}, \tag{8}$$

where $\boldsymbol{a}_i = [a_{i1}, a_{i2}, \dots, a_{in}]$, $i = 1, \dots, m$, and $\boldsymbol{b}_j = [b_{1j}, b_{2j}, \dots, b_{nj}]'$, $j = 1, \dots, p$. Then the product $\boldsymbol{C}_{m\times p} = \boldsymbol{A}_{m\times n}\boldsymbol{B}_{n\times p}$ is defined by the $m \times p$ matrix

$$\boldsymbol{C} = [c_{ij}]_{i=1,\dots,n;j=1,\dots,p} = [\langle \boldsymbol{a}_i, \boldsymbol{b}_j \rangle]_{i=1,\dots,n;j=1,\dots,p} = \left[\sum_{\ell=1}^{n} a_{i\ell} b_{\ell j}\right]_{i=1,\dots,n;j=1,\dots,p}. \tag{9}$$

For example, when

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 3 & 4 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 4 & 5 \\ 3 & 2 \\ 1 & 6 \end{bmatrix}, \tag{10}$$

then $\boldsymbol{C} = \boldsymbol{AB}$ is the 2×2 matrix

$$\begin{aligned} \boldsymbol{AB} &= \begin{bmatrix} 1 & 2 & 3 \\ 4 & 3 & 4 \end{bmatrix} \begin{bmatrix} 4 & 5 \\ 3 & 2 \\ 1 & 6 \end{bmatrix} = \begin{bmatrix} 1\times 4 + 2\times 3 + 3\times 1 & 1\times 5 + 2\times 2 + 3\times 6 \\ 4\times 4 + 3\times 3 + 4\times 1 & 4\times 5 + 3\times 2 + 4\times 6 \end{bmatrix} \\ &= \begin{bmatrix} 13 & 27 \\ 29 & 50 \end{bmatrix}. \end{aligned}$$

If column vectors $\boldsymbol{a} = [a_1, \dots, a_n]'$ and $\boldsymbol{b} = [b_1, \dots, b_m]'$, then, if $m = n$, the *inner product*

$$\boldsymbol{a}'\boldsymbol{b} = \sum_{i=1}^{n} a_i b_i, \tag{11}$$

and the outer product of $\boldsymbol{a}$ and $\boldsymbol{b}$ is the $n \times m$ matrix

$$\boldsymbol{a}\boldsymbol{b}' = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} [b_1, b_2, \dots, b_m] = \begin{bmatrix} a_1 b_1 & a_1 b_2 & \cdots & a_1 b_m \\ a_2 b_1 & a_2 b_2 & \cdots & a_2 b_m \\ \vdots & \vdots & \ddots & \vdots \\ a_n b_1 & a_n b_2 & \cdots & a_n b_m \end{bmatrix}. \tag{12}$$

The following rules are straightforward to verify:

(a) For conformable matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$, $\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C}) = \boldsymbol{AB}+\boldsymbol{AC}$ and $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$, but in general $\boldsymbol{AB} \neq \boldsymbol{BA}$.

(b) $(\boldsymbol{AB})' = \boldsymbol{B}'\boldsymbol{A}'$ For example, for the matrices in (10), we have

$$\boldsymbol{B}'\boldsymbol{A}' = \begin{bmatrix} 4 & 5 \\ 3 & 2 \\ 1 & 6 \end{bmatrix}' \begin{bmatrix} 1 & 2 & 3 \\ 4 & 3 & 4 \end{bmatrix}' = \begin{bmatrix} 4 & 3 & 1 \\ 5 & 2 & 6 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 2 & 3 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 13 & 29 \\ 27 & 50 \end{bmatrix} = (\boldsymbol{AB})'.$$

(c) For any $m \times n$ matrix $\boldsymbol{A}$, $\boldsymbol{I}_m\boldsymbol{A} = \boldsymbol{A}\boldsymbol{I}_n = \boldsymbol{A}$, where $\boldsymbol{I}_n$ is the identity of dimension n, as defined in (5).

The n nonzero vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ are *orthogonal* if $\boldsymbol{x}_i'\boldsymbol{x}_j = 0$ whenever $i \neq j$, and *orthonormal* if, in addition, $\boldsymbol{x}_i'\boldsymbol{x}_i = 1$for $i = 1, \ldots, n$. Clearly a set of orthogonal vectors can always be made orthonormal by scaling.
An $n \times n$ Matrix $\boldsymbol{A}$ is orthogonal if

$$\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I}, \tag{13}$$

i.e., if its columns and rows (viewed as vectors) are orthonormal.

2 Linear Independence and Rank

A set of n–dimensional vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$ is *linearly independent* if

$$\sum_{i=1}^{m} c_i \boldsymbol{x}_i = \boldsymbol{0} \tag{14}$$

implies $c_1 = c_2 = \cdots = c_m = 0$. Equivalently, the vectors are linearly independent if none can be written as a linear combination of the other vectors. Otherwise they are linearly dependent.

For example, vectors $\boldsymbol{e}_1 = [1, 0]'$ and $\boldsymbol{e}_2 = [0, 1]'$ are linearly independent, since

$$c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{15}$$

implies $c_1 = c_2 = 0$. It follows by the same argument that the columns (or rows) of the n–dimensional identity matrix are linearly independent, and they form the *canonical basis* of $\mathbb{R}^n$. As an example for linearly dependent vectors, consider $\boldsymbol{x}_1 = [1, 2]'$ and $\boldsymbol{x}_2 = [2, 4]'$. Here $\boldsymbol{x}_2 = 2\boldsymbol{x}_1$, so the linear combination (14) is zero for $c_1 = -2c_2$. Clearly two vectors are in general linearly dependent if one vector is a multiple of the other, since then $\boldsymbol{x}_1 = -(c_2/c_1)\boldsymbol{x}_2$.

An important property of a matrix is the number of linearly independent columns (which is equal to the number of linearly independent rows). This is the *rank* of an $m \times n$ matrix $\boldsymbol{A}$, denoted $r(\boldsymbol{A})$. Clearly $r(\boldsymbol{A}) \leq \min\{m, n\}$. Further properties of the rank:

R1: $r(\boldsymbol{AB}) \leq \min\{r(\boldsymbol{A}), r(\boldsymbol{B})\}$.

R2: $r(\boldsymbol{A}) = r(\boldsymbol{A}') = r(\boldsymbol{A}'\boldsymbol{A}) = r(\boldsymbol{A}\boldsymbol{A}')$.

2.1 Nonsingularity and Inverse

A square $n \times n$ matrix is *nonsingular* if it has *full rank*, i.e., its rank is n. Then (and only then) there exists a unique $n \times n$ matrix $\boldsymbol{B}$ such that $\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}$, which is the *inverse* of $\boldsymbol{A}$ and denoted as $\boldsymbol{A}^{-1}$. The solution of a system of n linear equations with n unknowns $\boldsymbol{Ax} = \boldsymbol{b}$ can then be written as $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$. The linear system has a unique solution if and only if $\boldsymbol{A}$ is nonsingular, and if and only if the system $\boldsymbol{Ax} = \boldsymbol{0}$ has only the solution $\boldsymbol{x} = \boldsymbol{0}$, i.e., for $\boldsymbol{Ax} = \boldsymbol{0}$ to have a nontrivial solution ($\boldsymbol{x} \neq \boldsymbol{0}$), we must have $\det(\boldsymbol{A}) = 0$.

For example, if

$$\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \tag{16}$$

then $\boldsymbol{A}$ is nonsingular if $\det(\boldsymbol{A}) = |\boldsymbol{A}| = ad - bc \neq 0$ and, in this case,

$$\boldsymbol{A}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \tag{17}$$

Useful properties:

N1: $(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$

N2: $(\boldsymbol{A}')^{-1} = (\boldsymbol{A}^{-1})'$, that is, the inverse of the transpose is the transpose of the inverse.

N3: Matrix $\boldsymbol{A}$ is nonsingular if and only if its determinant is different from zero.

N4: For $\boldsymbol{A}$ an $m \times n$ matrix and $\boldsymbol{P}$ and $\boldsymbol{Q}$ nonsingular $m \times m$ and $n \times n$ matrices, respectively, we have

$$r(\boldsymbol{A}) = r(\boldsymbol{PA}) = r(\boldsymbol{AQ}) = r(\boldsymbol{PAQ}). \tag{18}$$

N5: If matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are nonsingular and of the same order, $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$, which can easily be extended to more than two matrices, e.g., $(\boldsymbol{ABC})^{-1} = \boldsymbol{C}^{-1}(\boldsymbol{AB})^{-1} = \boldsymbol{C}^{-1}\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.

3 Trace

For an $n \times n$ matrix $\boldsymbol{A}$, the trace, denoted as $\mathrm{tr}(\boldsymbol{A})$, is the sum of the elements on the main diagonal, that is

$$\mathrm{tr}\boldsymbol{A} = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{i=1}^{n} a_{ii}. \tag{19}$$

Clearly $\mathrm{tr}(\alpha\boldsymbol{A}) = \alpha\mathrm{tr}\boldsymbol{A}$, and $\mathrm{tr}(\boldsymbol{A}+\boldsymbol{B}) = \mathrm{tr}\boldsymbol{A} + \mathrm{tr}\boldsymbol{B}$. Calculations also show that

$$\mathrm{tr}(\boldsymbol{AB}) = \mathrm{tr}(\boldsymbol{BA}), \tag{20}$$

which gives rise to the law of cyclical permutations, i.e., for conformable matrices,

$$\mathrm{tr}(\boldsymbol{ABC}) = \mathrm{tr}(\boldsymbol{CAB}) = \mathrm{tr}(\boldsymbol{BCA}). \tag{21}$$

4 Eigenvalues and Eigenvectors

An *eigenvalue* of an $n \times n$ matrix $\boldsymbol{A}$ is a (real or complex) scalar λ satisfying the equation

$$\boldsymbol{Ax} = \lambda\boldsymbol{x} \tag{22}$$

for some nonzero $n \times 1$ vector $\boldsymbol{x}$, which is called an eigenvector of $\boldsymbol{A}$. An eigenvector obviously is determined only up to a scalar multiple, since Equation (22) will also be satisfied for $\alpha\boldsymbol{x}$, $\alpha \neq 0$. Also, $\boldsymbol{x} \neq \boldsymbol{0}$ is assumed since (22) is satisfied for any λ if $\boldsymbol{x}$ is the zero vector.

Equation (22) can also be written as

$$(\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}, \tag{23}$$

which means that we require

$$\det(\lambda\boldsymbol{I} - \boldsymbol{A}) = 0, \tag{24}$$

which is a polynomial of degree n in λ, called the *characteristic equation* of matrix $\boldsymbol{A}$. Thus, an $n \times n$ matrix $\boldsymbol{A}$ will have n (real or complex) eigenvalues, counting multiplicities.

For example, for the 2×2 matrix (16), we have the characteristic equation

$$\begin{aligned}
\det(\lambda\boldsymbol{I} - \boldsymbol{A}) &= \det \begin{bmatrix} \lambda - a & -b \\ -c & \lambda - d \end{bmatrix} = (\lambda - a)(\lambda - d) - bc \\
&= \lambda^2 - (a + d)\lambda + ad - bc = \lambda^2 - \mathrm{tr}\boldsymbol{A}\lambda + \det(\boldsymbol{A}) = 0,
\end{aligned}$$

with solutions

$$\lambda_{1/2} = \frac{\mathrm{tr}\boldsymbol{A} \pm \sqrt{\mathrm{tr}\boldsymbol{A}^2 - 4\det\boldsymbol{A}}}{2}, \tag{25}$$

which shows that $\lambda_1 + \lambda_2 = \mathrm{tr}\boldsymbol{A}$ and $\lambda_1\lambda_2 = \det\boldsymbol{A}$. This is generally true, i.e., the sum of the eigenvalues of any square matrix is equal to its trace, and the product of the eigenvalues is equal to its determinant.

5 Symmetric Matrices

5.1 Quadratic Forms

Let $\boldsymbol{A}$ be a symmetric $n \times n$ matrix. The quadratic form associated with $\boldsymbol{A}$ is the function

$$\begin{aligned} f(\boldsymbol{x}) &= \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j a_{ij} = \sum_{i=1}^{n} x_i^2 a_{ii} + \sum_{i=1}^{n}\sum_{j\neq i} x_i x_j a_{ij} && (26) \\ &= \sum_{i=1}^{n} x_i^2 a_{ii} + 2\sum_{i}\sum_{j<i} x_i x_j a_{ij}, \quad \boldsymbol{x} \in \mathbb{R}^n. && (27) \end{aligned}$$

Clearly a quadratic form is always equal to zero if $\boldsymbol{x}$ is the zero vector.

Matrix $\boldsymbol{A}$ and the associated quadratic form are said to be

(a) *positive definite* if $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$ for any vector $\boldsymbol{x} \neq \mathbf{0}$, and

(b) *positive semidefinite* if $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0$ for any vector $\boldsymbol{x}$.

Negative definite and negative semidefinite matrices are analogously defined, and a matrix is indefinite if it is neither positive nor negative semidefinite.

(i) If $\boldsymbol{A}$ is a positive definite matrix, then it is nonsingular, and the inverse of $\boldsymbol{A}$ is also positive definite.

(ii) if $\boldsymbol{A}$ is $m \times n$, then $\boldsymbol{A}'\boldsymbol{A}$ is positive semidefinite. If $\boldsymbol{A}$ has rank n, then $\boldsymbol{A}'\boldsymbol{A}$ is positive definite.

(iii) Symmetric matrix is positive definite if and only if all eigenvalues are positive.

(iv) If $\boldsymbol{A}$ is positive definite, then $-\boldsymbol{A}$ is negative definite.

The 2×2 matrix

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \tag{28}$$

is positive definite if and only if

$$a_{11} > 0, \quad a_{11}a_{22} - a_{12}^2 > 0. \tag{29}$$

To see this, note that

$$\begin{aligned}
\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
&= x_1^2 a_1 1 + 2x_1 x_2 a_{12} + x_2^2 a_2 2 \\
&= a_{11}\left(x_1^2 + 2x_1 x_2 \frac{a_{12}}{a_{11}} + x_2^2 \frac{a_{12}^2}{a_{11}^2} \right) + x_2^2 a_{22} - x_2^2 \frac{a_{12}^2}{a_{11}} \\
&= a_{11}\left(x_1 + x_2 \frac{a_{12}}{a_{11}} \right)^2 + x_2^2 \frac{a_{11}a_{22} - a_{12}^2}{a_{11}}.
\end{aligned}$$

Example 1 *(Covariance matrices) Recall that the variance of a random variable X is defined by*

$$var(X) = E\{(X-\mu)^2\} = \int_{-\infty}^{\infty} (x-\mu_X)^2 f_X(x)dx, \tag{30}$$

where $\mu = \int x f_X(x)dx$ is the mean (or expectation) of X and f_X is the probability density function (pdf) of X. Now suppose we have a pair of random variables X_1 and X_2, then the covariance between X_1 and X_2, a measure of (linear) dependence, is defined by

$$cov(X_1, X_2) = E[(X_1-\mu_1)(X_2-\mu_2)] = E(X_1X_2) - \mu_1\mu_2. \tag{31}$$

Now consider random vector $\boldsymbol{x} = [X_1, X_2, \ldots, X_n]'$. The expectation $E\boldsymbol{x}$ is just the column vector $[\mu_1, \ldots, \mu_n]'$, where $\mu_i = E(X_i)$, $i = 1, \ldots .n$. The covariance matrix (or variance–covariance matrix) of the random vector $\boldsymbol{x}$ is

$$\begin{aligned}
\boldsymbol{V} = cov(\boldsymbol{x}) &= E[(\boldsymbol{x} - E\boldsymbol{x})(\boldsymbol{x} - E\boldsymbol{x})'] && (32) \\
&= E\left\{ \begin{bmatrix} X_1-\mu_1 \\ X_2-\mu_2 \\ \vdots \\ X_n-\mu_n \end{bmatrix} \begin{bmatrix} X_1-\mu_1 & X_2-\mu_2 & \cdots & X_n-\mu_n \end{bmatrix} \right\} && (33) \\
&= \begin{bmatrix} E(X_1-\mu_1)^2 & E(X_1-\mu_1)(X_2-\mu_2) & \cdots & E(X_1-\mu_1)(X_n-\mu_n) \\ E(X_2-\mu_2)(X_1-\mu_1) & E(X_2-\mu_2)^2 & \cdots & E(X_2-\mu_2)(X_n-\mu_n) \\ \vdots & \vdots & \ddots & \vdots \\ E(X_n-\mu_n)(X_1-\mu_1) & E(X_n-\mu_n)(X_2-\mu_2) & \cdots & E(X_n-\mu_n)^2 \end{bmatrix} && (34) \\
&= \begin{bmatrix} var(X_1) & cov(X_1,X_2) & \cdots & cov(X_1,X_n) \\ cov(X_1,X_2) & var(X_2) & \cdots & cov(X_2,X_n) \\ \vdots & \vdots & \ddots & \vdots \\ cov(X_1,X_n) & cov(X_2,X_n) & \cdots & var(X_n) \end{bmatrix} && (35)
\end{aligned}$$

Recall that a linear combination of random variables, such as

$$Y = \boldsymbol{a}'\boldsymbol{x} = \sum_{i=1}^{n} a_i X_i \tag{36}$$

has expectation

$$E(Y) = E\left\{\sum_{i=1}^{n} a_i X_i\right\} = \sum_{i=1}^{n} a_i E(X_i) = \boldsymbol{a}' E\boldsymbol{x}. \tag{37}$$

The variance of Y *defined by (36) is*

$$\begin{aligned}
0 \le var(Y) &= E\big\{(Y - EY)^2\big\} = E\left\{\left(\sum_{i=1}^{n} a_i (X_i - \mu_i)\right)^2\right\} && (38)\\
&= E\left\{\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (X_i - \mu_i)(X_j - \mu_j)\right\} && (39)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_j a_j E(X_i - \mu_i)(X_j - \mu_j) && (40)\\
&= \sum_{i=1}^{n} a_i^2 var(X_i) + \sum_{i}\sum_{j\neq i} a_i a_j cov(X_i, X_j) && (41)\\
&= \boldsymbol{a}'\boldsymbol{V}\boldsymbol{a}. && (42)
\end{aligned}$$

This shows that the variance of a linear combination of random variables can be written as a quadratic form in the covariance matrix of these variables, and since a variance is nonnegative (positive if the variable is not a constant), we see that covariance matrices are always positive semidefinite, *and they are positive definite unless there exists* $\boldsymbol{a} \neq \boldsymbol{0}$ *such that the linear combination* $\boldsymbol{a}'\boldsymbol{x}$ *is a constant.*

For two random variables X *and* Y*, the covariance matrix is*

$$\boldsymbol{V} = \begin{bmatrix} var(X) & cov(X,Y) \\ cov(X,Y) & var(Y) \end{bmatrix} =: \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix}. \tag{43}$$

From (29) and the positive (semi)definiteness of covariance matrices, it follows that

$$\sigma_X \sigma_Y = \sqrt{\sigma_X^2 \sigma_Y^2} \ge |\sigma_{XY}|, \tag{44}$$

which among other things implies that the correlation coefficient

$$corr(X,Y) = \rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \in [-1, 1]. \tag{45}$$

$|\rho| = 1$ *if and only if the variables are linearly dependent, i.e., each can be written as a linear function of the other.*

5.2 Choleski Decomposition of Positive Definite Matrices

Suppose random variable X has mean μ and variance σ^2. Then by standardizing X we get a variable with mean zero and unit variance,

$$Z = \frac{X - \mu}{\sigma}. \tag{46}$$

The same can be done for a random vector $\boldsymbol{x}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{V}$.

This is based on the following result: *(Choleski Decomposition) If $\boldsymbol{A}$ is an $n \times n$ positive definite symmetric matrix, there is a nonsingular lower triangular matrix $\boldsymbol{C}$ with positive diagonal elements such that $\boldsymbol{CC'}$.*

Note that $\boldsymbol{C}$ may be viewed as the "matrix root" of $\boldsymbol{A}$. As a simple example, consider the matrix

$$\begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 1 & \sqrt{3} \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & \sqrt{3} \end{bmatrix}. \tag{47}$$

Now let $\boldsymbol{V} = \boldsymbol{CC'}$, so that $\boldsymbol{V}^{-1} = \boldsymbol{C}'^{-1}\boldsymbol{C}^{-1}$. Consider the random vector $\boldsymbol{z} = \boldsymbol{C}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$, which has zero mean vector, so the covariance matrix of $\boldsymbol{z}$ is

$$\begin{aligned} \mathrm{cov}(\boldsymbol{z}) &= \mathrm{E}(\boldsymbol{zz'}) = \mathrm{E}\{\boldsymbol{C}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{C}'^{-1}\} && (48) \\ &= \boldsymbol{C}^{-1}\mathrm{E}\{(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})'\}\boldsymbol{C}'^{-1} && (49) \\ &= \boldsymbol{C}^{-1}\boldsymbol{V}\boldsymbol{C}'^{-1} && (50) \\ &= \boldsymbol{C}^{-1}\boldsymbol{CC'}\boldsymbol{C}'^{-1} && (51) \\ &= \boldsymbol{I}, && (52) \end{aligned}$$

the identity matrix, i.e., the variables are uncorrelated with unit variance.

In general, if random vector $\boldsymbol{x}$ is an $n \times 1$ random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{V}$, then the vector $\boldsymbol{y} = \boldsymbol{Ax}+\boldsymbol{b}$ has mean vector $\boldsymbol{A\mu}+\boldsymbol{b}$ and covariance matrix $\boldsymbol{AVA'}$, the calculation of this being basically the same as in (48).

Moreover, if $\boldsymbol{x}$ has a multivariate normal distribution, written $\boldsymbol{x} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{V})$, then, as linear combinations of normally distributed random variables are likewise normally distributed, we have

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} \sim \mathrm{N}(\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}, \boldsymbol{A}\boldsymbol{V}\boldsymbol{A}'). \tag{53}$$

6 Vector and Matrix Differentiation

For function

$$y = f(\boldsymbol{x}) = f(x_1, \ldots, x_n), \tag{54}$$

the gradient is the column vector

$$\frac{\partial f}{\partial \boldsymbol{x}} = \left[\begin{array}{c} \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \end{array} \right]', \tag{55}$$

and the (symmetric under mild conditions) Hessian matrix of second derivatives

$$\boldsymbol{H} = \frac{\partial^2 f}{\partial \boldsymbol{x} \partial \boldsymbol{x}'} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \tag{56}$$

We are interest in derivatives of linear and quadratic forms. If

$$f(\boldsymbol{x}) = \boldsymbol{a}'\boldsymbol{x} = \sum_{i=1}^{n} a_i x_i, \tag{57}$$

then

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{a}. \tag{58}$$

If

$$f(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j, \tag{59}$$

we observe that for general $\boldsymbol{A}$

$$\begin{aligned} \frac{\partial f(\boldsymbol{x})}{\partial x_\ell} &= 2a_{\ell\ell}x_\ell + \sum_{i \neq \ell} a_{i\ell}x_i + \sum_{j \neq \ell} a_{\ell j}x_j & (60) \\ &= \sum_{i=1}^{n} a_{i\ell}x_i + \sum_{j=1}^{n} a_{\ell j}x_j & (61) \end{aligned}$$

which can be identified as the ℓth element of the vector $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}'\boldsymbol{x} = (\boldsymbol{A} + \boldsymbol{A}')\boldsymbol{x}$; hence

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = (\boldsymbol{A} + \boldsymbol{A}')\boldsymbol{x}. \tag{62}$$

The case of a symmetric $\boldsymbol{A}$ is of particular interest in our applications. Since a square matrix $\boldsymbol{A}$ is symmetric exactly if $\boldsymbol{A} = \boldsymbol{A}'$, we have

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} = 2\boldsymbol{A}\boldsymbol{x}, \tag{63}$$

in this case, and the Hessian is

$$\frac{\partial^2 f}{\partial \boldsymbol{x} \partial \boldsymbol{x}'} = 2\boldsymbol{A}. \tag{64}$$

For a function of the (bilinear) form

$$f(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{y}'\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{A}'\boldsymbol{y}, \tag{65}$$

we have from (58) that

$$\frac{\partial f}{\partial \boldsymbol{x}} = \boldsymbol{A}'\boldsymbol{y}, \tag{66}$$

and

$$\frac{\partial f}{\partial \boldsymbol{y}} = \boldsymbol{A}\boldsymbol{x}. \tag{67}$$

For a function $f(\boldsymbol{x})$ to have a local minimum or maximum at a point $\boldsymbol{x}_0$, its gradient must vanish in this point. Moreover, such a point is

(i) a local maximum if the Hessian evaluated at this point is negative definite,

(ii) a local minimum if the Hessian evaluated at this point is positive definite.