2.14 VECTOR AND MATRIX CALCULUS

2.14.1 Derivatives of Functions of Vectors and Matrices

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

$$\frac{\partial u}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial u}{\partial x_1} \\ \frac{\partial u}{\partial x_2} \\ \vdots \\ \frac{\partial u}{\partial x_p} \end{pmatrix}. \tag{2.111}$$

Two specific functions of interest are $u = \mathbf{a}'\mathbf{x}$ and $u = \mathbf{x}'\mathbf{A}\mathbf{x}$. Their derivatives with respect to $\mathbf{x}$ are given in the following two theorems.

Theorem 2.14a. Let $u = \mathbf{a}'\mathbf{x} = \mathbf{x}'\mathbf{a}$, where $\mathbf{a}' = (a_1, a_2, \ldots, a_p)$ is a vector of constants. Then

$$\frac{\partial u}{\partial \mathbf{x}} = \frac{\partial(\mathbf{a}'\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial(\mathbf{x}'\mathbf{a})}{\partial \mathbf{x}} = \mathbf{a}. \tag{2.112}$$

PROOF

$$\frac{\partial u}{\partial x_i} = \frac{\partial(a_1x_1 + a_2x_2 + \cdots + a_px_p)}{\partial x_i} = a_i.$$

Thus by (2.111) we obtain

$$\frac{\partial u}{\partial \mathbf{x}} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} = \mathbf{a}.$$

□

Theorem 2.14b. Let $u = \mathbf{x}'\mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is a symmetric matrix of constants. Then

$$\frac{\partial u}{\partial \mathbf{x}} = \frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}. \tag{2.113}$$

PROOF. We demonstrate that (2.113) holds for the special case in which $\mathbf{A}$ is 3×3. The illustration could be generalized to a symmetric $\mathbf{A}$ of any size. Let

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad \text{and} \quad \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix} = \begin{pmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \mathbf{a}_3' \end{pmatrix}.$$

Then $\mathbf{x}'\mathbf{A}\mathbf{x} = x_1^2 a_{11} + 2x_1x_2a_{12} + 2x_1x_3a_{13} + x_2^2a_{22} + 2x_2x_3a_{23} + x_3^2a_{33}$, and we have

$$\frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial x_1} = 2x_1a_{11} + 2x_2a_{12} + 2x_3a_{13} = 2\mathbf{a}_1'\mathbf{x}$$

$$\frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial x_2} = 2x_1a_{12} + 2x_2a_{22} + 2x_3a_{23} = 2\mathbf{a}_2'\mathbf{x}$$

$$\frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial x_3} = 2x_1a_{13} + 2x_2a_{23} + 2x_3a_{33} = 2\mathbf{a}_3'\mathbf{x}.$$

Thus by (2.11), (2.27), and (2.111), we obtain

$$\frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial \mathbf{x}} = \begin{pmatrix} \dfrac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial x_1} \\ \dfrac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial x_2} \\ \dfrac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial x_3} \end{pmatrix} = 2\begin{pmatrix} \mathbf{a}_1'\mathbf{x} \\ \mathbf{a}_2'\mathbf{x} \\ \mathbf{a}_3'\mathbf{x} \end{pmatrix} = 2\mathbf{A}\mathbf{x}.$$

□

Now let $u = f(\mathbf{X})$ be a function of the variables $x_{11}, x_{12}, \ldots, x_{pp}$ in the $p \times p$ matrix $\mathbf{X}$, and let $(\partial u/\partial x_{11}), (\partial u/\partial x_{12}), \ldots, (\partial u/\partial x_{pp})$ be the partial derivatives. Similarly to (2.111), we define $\partial u/\partial \mathbf{X}$ as

$$\frac{\partial u}{\partial \mathbf{X}} = \begin{pmatrix} \dfrac{\partial u}{\partial x_{11}} & \cdots & \dfrac{\partial u}{\partial x_{1p}} \\ \vdots & & \vdots \\ \dfrac{\partial u}{\partial x_{p1}} & \cdots & \dfrac{\partial u}{\partial x_{pp}} \end{pmatrix}. \tag{2.114}$$

Two functions of interest of this type are $u = \text{tr}(\mathbf{XA})$ and $u = \ln|\mathbf{X}|$ for a positive definite matrix $\mathbf{X}$.

Theorem 2.14c. Let $u = \text{tr}(\mathbf{XA})$, where $\mathbf{X}$ is a $p \times p$ positive definite matrix and $\mathbf{A}$ is a $p \times p$ matrix of constants. Then

$$\frac{\partial u}{\partial \mathbf{X}} = \frac{\partial[\text{tr}(\mathbf{XA})]}{\partial \mathbf{X}} = \mathbf{A} + \mathbf{A}' - \text{diag}\,\mathbf{A}. \tag{2.115}$$

PROOF. Note that $\mathrm{tr}(\mathbf{XA}) = \sum_{i=1}^{p}\sum_{j=1}^{p} x_{ij}a_{ji}$ [see the proof of Theorem 2.11(ii)]. Since $x_{ij} = x_{ji}$, $[\partial \mathrm{tr}(\mathbf{XA})]/\partial x_{ij} = a_{ji} + a_{ij}$ if $i \neq j$, and $[\partial \mathrm{tr}(\mathbf{XA})]/\partial x_{ii} = a_{ii}$. The result follows. □

Theorem 2.14d. Let $u = \ln|\mathbf{X}|$ where $\mathbf{X}$ is a $p \times p$ positive definite matrix. Then

$$\frac{\partial \ln|\mathbf{X}|}{\partial \mathbf{X}} = 2\mathbf{X}^{-1} - \mathrm{diag}(\mathbf{X}^{-1}). \tag{2.116}$$

PROOF. See Harville (1997, p. 306). See Problem 2.83 for a demonstration that this theorem holds for 2×2 matrices. □

2.14.2 Derivatives Involving Inverse Matrices and Determinants

Let $\mathbf{A}$ be an $n \times n$ nonsingular matrix with elements a_{ij} that are functions of a scalar x. We define $\partial\mathbf{A}/\partial x$ as the $n \times n$ matrix with elements $\partial a_{ij}/\partial x$. The related derivative $\partial\mathbf{A}^{-1}/\partial x$ is often of interest. If $\mathbf{A}$ is positive definite, the derivative $(\partial/\partial x)\log|\mathbf{A}|$ is also often of interest.

Theorem 2.14e. Let $\mathbf{A}$ be nonsingular of order n with derivative $\partial\mathbf{A}/\partial x$. Then

$$\frac{\partial \mathbf{A}^{-1}}{\partial x} = -\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial x}\mathbf{A}^{-1} \tag{2.117}$$

PROOF. Because $\mathbf{A}$ is nonsingular, we have

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

Thus

$$\frac{\partial \mathbf{A}^{-1}}{\partial x}\mathbf{A} + \mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial x} = \mathbf{O}.$$

Hence

$$\frac{\partial \mathbf{A}^{-1}}{\partial x}\mathbf{A} = -\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial x},$$

and so

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

□

Theorem 2.14f. Let $\mathbf{A}$ be an $n \times n$ positive define matrix. Then

$$\frac{\partial \log|\mathbf{A}|}{\partial x} = \mathrm{tr}\left(\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial x}\right). \tag{2.118}$$

PROOF. Since $\mathbf{A}$ is positive definite, its spectral decomposition (Theorem 2.12d) can be written as $\mathbf{CDC}'$, where $\mathbf{C}$ is an orthogonal matrix and $\mathbf{D}$ is a diagonal matrix of positive eigenvalues, λ_i. Using Theorem 2.12e, we obtain

$$\begin{aligned}\frac{\partial \log|\mathbf{A}|}{\partial x} &= \frac{\partial \log \prod_{i=1}^{n} \lambda_i}{\partial x} \\ &= \frac{\partial \sum_{i=1}^{n} \log \lambda_i}{\partial x} \\ &= \sum_{i=1}^{n} \frac{1}{\lambda_i} \frac{\partial \lambda_i}{\partial x} \\ &= \mathrm{tr}\left(\mathbf{D}^{-1} \frac{\partial \mathbf{D}}{\partial x}\right).\end{aligned}$$

Now

$$\begin{aligned}\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial x} &= \mathbf{CD}^{-1}\mathbf{C}' \frac{\partial \mathbf{CDC}'}{\partial x} \\ &= \mathbf{CD}^{-1}\mathbf{C}' \left[\mathbf{C} \frac{\partial \mathbf{DC}'}{\partial x} + \frac{\partial \mathbf{C}}{\partial x} \mathbf{DC}'\right] \\ &= \mathbf{CD}^{-1}\mathbf{C}' \left[\mathbf{C} \frac{\partial \mathbf{D}}{\partial x} \mathbf{C}' + \mathbf{CD} \frac{\partial \mathbf{C}'}{\partial x} + \frac{\partial \mathbf{C}}{\partial x} \mathbf{DC}'\right] \\ &= \mathbf{CD}^{-1} \frac{\partial \mathbf{D}}{\partial x} \mathbf{C}' + \mathbf{C} \frac{\partial \mathbf{C}'}{\partial x} + \mathbf{CD}^{-1}\mathbf{C}' \frac{\partial \mathbf{C}}{\partial x} \mathbf{DC}'.\end{aligned}$$

Using Theorem 2.11(i) and (ii), we have

$$\mathrm{tr}\left(\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial x}\right) = \mathrm{tr}\left(\mathbf{D}^{-1} \frac{\partial \mathbf{D}}{\partial x} + \mathbf{C} \frac{\partial \mathbf{C}'}{\partial x} + \mathbf{C}' \frac{\partial \mathbf{C}}{\partial x}\right).$$

Since $\mathbf{C}$ is orthogonal, $\mathbf{C}'\mathbf{C} = \mathbf{I}$ which implies that

$$\frac{\partial \mathbf{C}'\mathbf{C}}{\partial x} = \mathbf{C}' \frac{\partial \mathbf{C}}{\partial x} + \frac{\partial \mathbf{C}'}{\partial x} \mathbf{C} = \mathbf{O}$$

and

$$\mathrm{tr}\left(\mathbf{C}' \frac{\partial \mathbf{C}}{\partial x} + \frac{\partial \mathbf{C}'\mathbf{C}}{\partial x}\right) = \mathrm{tr}\left(\mathbf{C}' \frac{\partial \mathbf{C}}{\partial x} + \mathbf{C} \frac{\partial \mathbf{C}'}{\partial x}\right) = 0.$$

Thus $\mathrm{tr}[\mathbf{A}^{-1}(\partial \mathbf{A}/\partial x)] = \mathrm{tr}[\mathbf{D}^{-1}(\partial \mathbf{D}/\partial x)]$ and the result follows. □

2.14.3 Maximization or Minimization of a Function of a Vector

Consider a function $u = f(\mathbf{x})$ of the p variables in $\mathbf{x}$. In many cases we can find a maximum or minimum of u by solving the system of p equations

$$\frac{\partial u}{\partial \mathbf{x}} = \mathbf{0}. \tag{2.119}$$

Occasionally the situation requires the maximization or minimization of the function u, subject to q constraints on $\mathbf{x}$. We denote the constraints as $h_1(\mathbf{x}) = 0,\ h_2(\mathbf{x}) = 0, \ldots, h_q(\mathbf{x}) = 0$ or, more succinctly, $\mathbf{h}(\mathbf{x}) = \mathbf{0}$. Maximization or minimization of u subject to $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ can often be carried out by the method of Lagrange multipliers. We denote a vector of q unknown constants (the *Lagrange multipliers*) by $\boldsymbol{\lambda}$ and let $\mathbf{y}' = (\mathbf{x}', \boldsymbol{\lambda}')$. We then let $v = u + \boldsymbol{\lambda}'\mathbf{h}(\mathbf{x})$. The maximum or minimum of u subject to $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ is obtained by solving the equations

$$\frac{\partial v}{\partial \mathbf{y}} = \mathbf{0}$$

or, equivalently

$$\frac{\partial u}{\partial \mathbf{x}} + \frac{\partial \mathbf{h}}{\partial \mathbf{x}}\boldsymbol{\lambda} = \mathbf{0} \quad \text{and} \quad \mathbf{h}(\mathbf{x}) = \mathbf{0}, \tag{2.120}$$

where

$$\frac{\partial \mathbf{h}}{\partial \mathbf{x}} = \begin{pmatrix} \dfrac{\partial h_1}{\partial x_1} & \cdots & \dfrac{\partial h_q}{\partial x_1} \\ \vdots & & \vdots \\ \dfrac{\partial h_1}{\partial x_p} & \cdots & \dfrac{\partial h_q}{\partial x_p} \end{pmatrix}.$$

PROBLEMS

2.1 Prove Theorem 2.2a.

2.2 Let $\mathbf{A} = \begin{pmatrix} 7 & -3 & 2 \\ 4 & 9 & 5 \end{pmatrix}$.

(**a**) Find $\mathbf{A}'$.

(**b**) Verify that $(\mathbf{A}')' = \mathbf{A}$, thus illustrating Theorem 2.1.

(**c**) Find $\mathbf{A}'\mathbf{A}$ and $\mathbf{A}\mathbf{A}'$.

2.3 Let $\mathbf{A} = \begin{pmatrix} 2 & 4 \\ -1 & 3 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 3 \\ 2 & -1 \end{pmatrix}$.