Appendix A
Differentiation of Matrix Expressions

For univariate x and $f : \mathbb{R} \rightarrow \mathbb{R}$ we write the derivative as

$$\frac{df}{dx}.$$

We define

$$\frac{\partial}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_p} \end{bmatrix}$$

to be differentiation with respect to elements of a vector $\boldsymbol{x} = [x_1, \dots, x_p]^{\mathrm{T}}$. Let $\boldsymbol{a}$ and $\boldsymbol{x}$ represent $p \times 1$ vectors, then

$$\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}) = \boldsymbol{a} = \frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{a}), \tag{A.1}$$

the second equality arising because $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{a}$. Also

$$\frac{\partial}{\partial \boldsymbol{x}^{\mathrm{T}}}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}) = \left[\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x})\right]^{\mathrm{T}} = \boldsymbol{a}^{\mathrm{T}} = \frac{\partial}{\partial \boldsymbol{x}^{\mathrm{T}}}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{a}). \tag{A.2}$$

Suppose $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{x})$ is an $r \times 1$ vector and $\boldsymbol{x}$ is $p \times 1$. Then

$$\frac{\partial \boldsymbol{u}^{\mathrm{T}}}{\partial \boldsymbol{x}}$$

is a matrix of order $p \times r$ with (i, j)th element

$$\frac{\partial u_j}{\partial x_i}, \quad i = 1, \dots, p, \quad j = 1, \dots, r.$$

J. Wakefield, *Bayesian and Frequentist Regression Methods*, Springer Series in Statistics, DOI 10.1007/978-1-4419-0925-1_13,
© Springer Science+Business Media New York 2013

The transpose

$$\left(\frac{\partial \boldsymbol{u}^{\mathrm{T}}}{\partial \boldsymbol{x}}\right)^{\mathrm{T}} = \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{x}^{\mathrm{T}}}$$

is a matrix of order $r \times p$ with (j,k)th element

$$\frac{\partial u_j}{\partial x_k}, \quad j = 1, \ldots, r, \quad k = 1, \ldots, p.$$

For example,

$$\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{x}^{\mathrm{T}}} = \frac{\partial \boldsymbol{x}^{\mathrm{T}}}{\partial \boldsymbol{x}} = \mathbf{I}_p,$$

the $p \times p$ identity matrix.

Consider the matrix $\boldsymbol{A}$ of dimension $p \times p$. If $\boldsymbol{A}$ is not a function of $\boldsymbol{x}$:

$$\frac{\partial}{\partial \boldsymbol{x}^{\mathrm{T}}}(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A}\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{x}^{\mathrm{T}}} = \boldsymbol{A}$$

and

$$\frac{\partial}{\partial \boldsymbol{x}^{\mathrm{T}}}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}) = \frac{\partial \boldsymbol{x}^{\mathrm{T}}}{\partial \boldsymbol{x}}\boldsymbol{A} = \boldsymbol{A}.$$

If $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{x})$ then

$$\frac{\partial}{\partial \boldsymbol{x}^{\mathrm{T}}}(\boldsymbol{A}\boldsymbol{u}) = \boldsymbol{A}\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{x}^{\mathrm{T}}},$$

and

$$\frac{\partial}{\partial \boldsymbol{x}^{\mathrm{T}}}(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}) = \frac{\partial \boldsymbol{u}^{\mathrm{T}}}{\partial \boldsymbol{x}}\boldsymbol{A}.$$

Let $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{x})$ and $\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{x})$ be $p \times 1$ vectors. Then the derivative of the inner product $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}$ is

$$\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}) = \frac{\partial \boldsymbol{u}^{\mathrm{T}}}{\partial \boldsymbol{x}}\boldsymbol{v} + \frac{\partial \boldsymbol{v}^{\mathrm{T}}}{\partial \boldsymbol{x}}\boldsymbol{u}.$$

If $\boldsymbol{A}$ is again a $p \times p$ matrix then

$$\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u}) = \frac{\partial \boldsymbol{u}^{\mathrm{T}}}{\partial \boldsymbol{x}}\boldsymbol{A}\boldsymbol{u} + \frac{\partial \boldsymbol{v}^{\mathrm{T}}}{\partial \boldsymbol{x}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}.$$

If $\boldsymbol{A}$ is symmetric

$$\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u}) = 2\frac{\partial \boldsymbol{u}^{\mathrm{T}}}{\partial \boldsymbol{x}}\boldsymbol{A}\boldsymbol{u}.$$

In particular, for a quadratic form

$$\frac{\partial}{\partial \boldsymbol{x}}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x}. \tag{A.3}$$

Let $f : \mathbb{R} \to \mathbb{R}$ then

$$\frac{\partial f}{\partial \boldsymbol{x}}$$

is $p \times 1$ and

$$\frac{\partial}{\partial \boldsymbol{x}} \frac{\partial f}{\partial \boldsymbol{x}^{\mathrm{T}}} = \frac{\partial^2 f}{\partial \boldsymbol{x} \partial \boldsymbol{x}^{\mathrm{T}}}$$

is a $p \times p$ matrix with elements

$$\frac{\partial^2 f}{\partial x_i \partial x_j}, \quad i = 1, \ldots, p, \;\; j = 1, \ldots, p.$$

For example, with $p = 2$:

$$\frac{\partial}{\partial \boldsymbol{x}} \frac{\partial f}{\partial \boldsymbol{x}^{\mathrm{T}}} = \begin{bmatrix} \frac{\partial}{\partial x_1} \begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{pmatrix} \\ \frac{\partial}{\partial x_2} \begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \end{pmatrix} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}.$$

For a non-singular $p \times p$ matrix $\boldsymbol{A}$, whose elements are functions of x, we have

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

Also,

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

The trace of a $p \times p$ square matrix is

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