Matrix Differentiation

(and some other stuff)

Randal J. Barnes
Department of Civil Engineering, University of Minnesota
Minnesota, USA

1 Introduction

Throughout this presentation I have chosen to use a *symbolic matrix notation*. This choice was not made lightly. I am a strong advocate of index notation, when appropriate. For example, index notation greatly simplifies the presentation and manipulation of differential geometry. As a rule-of-thumb, if your work is going to primarily involve differentiation with respect to the spatial coordinates, then index notation is almost surely the appropriate choice.

In the present case, however, I will be manipulating large systems of equations in which the matrix calculus is relatively simply while the matrix algebra and matrix arithmetic is messy and more involved. Thus, I have chosen to use symbolic notation.

2 Notation and Nomenclature

Definition 1 Let $a_{ij} \in \mathfrak{R}$, i = 1, 2, ..., m, j = 1, 2, ..., n. Then the ordered rectangular array

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

is said to be a real *matrix* of dimension $m \times n$.

When writing a matrix I will occasionally write down its typical element as well as its dimension. Thus,

$$\mathbf{A} = [a_{ij}], \quad i = 1, 2, \dots, m; \ j = 1, 2, \dots, n,$$
 (2)

denotes a matrix with \mathfrak{m} rows and \mathfrak{n} columns, whose typical element is \mathfrak{a}_{ij} . Note, the first subscript locates the *row* in which the typical element lies while the second subscript locates the *column*. For example, \mathfrak{a}_{jk} denotes the element lying in the jth row and kth column of the matrix \mathbf{A} .

Definition 2 A *vector* is a matrix with only one column. Thus, all vectors are inherently column vectors.

Convention 1

Multi-column matrices are denoted by boldface uppercase letters: for example, $\mathbf{A}, \mathbf{B}, \mathbf{X}$. Vectors (single-column matrices) are denoted by boldfaced lowercase letters: for example, $\mathbf{a}, \mathbf{b}, \mathbf{x}$. I will attempt to use letters from the beginning of the alphabet to designate known matrices, and letters from the end of the alphabet for unknown or variable matrices.

Convention 2

When it is useful to explicitly attach the matrix dimensions to the symbolic notation, I will use an underscript. For example, $\underset{m \times n}{\textbf{A}}$, indicates a known, multi-column matrix with m rows and n columns.

A superscript ^T denotes the matrix transpose operation; for example, \mathbf{A}^T denotes the transpose of \mathbf{A} . Similarly, if \mathbf{A} has an inverse it will be denoted by \mathbf{A}^{-1} . The determinant of \mathbf{A} will be denoted by either $|\mathbf{A}|$ or $\det(\mathbf{A})$. Similarly, the rank of a matrix \mathbf{A} is denoted by rank(\mathbf{A}). An identity matrix will be denoted by \mathbf{I} , and $\mathbf{0}$ will denote a null matrix.

3 Matrix Multiplication

Definition 3 Let **A** be $m \times n$, and **B** be $n \times p$, and let the product **AB** be

$$C = AB \tag{3}$$

then **C** is a $m \times p$ matrix, with element (i, j) given by

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \tag{4}$$

for all i = 1, 2, ..., m, j = 1, 2, ..., p.

Proposition 1 Let A be $m \times n$, and x be $n \times 1$, then the typical element of the product

$$z = Ax \tag{5}$$

is given by

$$z_{i} = \sum_{k=1}^{n} a_{ik} x_{k} \tag{6}$$

for all i = 1, 2, ..., m. Similarly, let **y** be $m \times 1$, then the typical element of the product

$$\mathbf{z}^{\mathsf{T}} = \mathbf{y}^{\mathsf{T}} \mathbf{A} \tag{7}$$

is given by

$$z_{i} = \sum_{k=1}^{n} \alpha_{ki} y_{k} \tag{8}$$

for all i = 1, 2, ..., n. Finally, the scalar resulting from the product

$$\alpha = \mathbf{y}^{\mathsf{T}} \mathbf{A} \mathbf{x} \tag{9}$$

is given by

$$\alpha = \sum_{j=1}^{m} \sum_{k=1}^{n} \alpha_{jk} y_j x_k \tag{10}$$

Proof: These are merely direct applications of Definition 3. q.e.d.

Proposition 2 Let **A** be $m \times n$, and **B** be $n \times p$, and let the product **AB** be

$$\mathbf{C} = \mathbf{A}\mathbf{B} \tag{11}$$

then

$$\mathbf{C}^{\mathsf{T}} = \mathbf{B}^{\mathsf{T}} \mathbf{A}^{\mathsf{T}} \tag{12}$$

Proof: The typical element of **C** is given by

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \tag{13}$$

By definition, the typical element of C^T , say d_{ij} , is given by

$$d_{ij} = c_{ji} = \sum_{k=1}^{n} a_{jk} b_{ki}$$

$$\tag{14}$$

Hence,

$$\mathbf{C}^{\mathsf{T}} = \mathbf{B}^{\mathsf{T}} \mathbf{A}^{\mathsf{T}} \tag{15}$$

q.e.d.

Proposition 3 Let A and B be $n \times n$ and invertible matrices. Let the product AB be given by

$$C = AB \tag{16}$$

then

$$\mathbf{C}^{-1} = \mathbf{B}^{-1} \mathbf{A}^{-1} \tag{17}$$

Proof:

$$\mathbf{C}\mathbf{B}^{-1}\mathbf{A}^{-1} = \mathbf{A}\mathbf{B}\mathbf{B}^{-1}\mathbf{A}^{-1} = \mathbf{I} \tag{18}$$

q.e.d.

4 Partioned Matrices

Frequently, I will find it convenient to deal with *partitioned matrices* ¹. Such a representation, and the manipulation of this representation, are two of the relative advantages of the symbolic matrix notation.

Definition 4 Let **A** be $m \times n$ and write

$$\mathbf{A} = \begin{bmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{D} & \mathbf{E} \end{bmatrix} \tag{19}$$

where **B** is $m_1 \times n_1$, **E** is $m_2 \times n_2$, **C** is $m_1 \times n_2$, **D** is $m_2 \times n_1$, $m_1 + m_2 = m$, and $n_1 + n_2 = n$. The above is said to be a *partition* of the matrix **A**.

¹Much of the material in this section is extracted directly from Dhrymes (1978, Section 2.7).

Proposition 4 Let **A** be a square, nonsingular matrix of order **m**. Partition **A** as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \tag{20}$$

so that \mathbf{A}_{11} is a nonsingular matrix of order \mathbf{m}_1 , \mathbf{A}_{22} is a nonsingular matrix of order \mathbf{m}_2 , and $\mathbf{m}_1 + \mathbf{m}_2 = \mathbf{m}$. Then

$$\mathbf{A}^{-1} = \begin{bmatrix} \left(\mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21} \right)^{-1} & -\mathbf{A}_{11}^{-1} \mathbf{A}_{12} \left(\mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12} \right)^{-1} \\ -\mathbf{A}_{22}^{-1} \mathbf{A}_{21} \left(\mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21} \right)^{-1} & \left(\mathbf{A}_{22} - \mathbf{A}_{21} \mathbf{A}_{11}^{-1} \mathbf{A}_{12} \right)^{-1} \end{bmatrix}$$
(21)

Proof: Direct multiplication of the proposed A^{-1} and A yields

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I} \tag{22}$$

q.e.d.

5 Matrix Differentiation

In the following discussion I will differentiate matrix quantities with respect to the elements of the referenced matrices. Although no new concept is required to carry out such operations, the element-by-element calculations involve cumbersome manipulations and, thus, it is useful to derive the necessary results and have them readily available ².

Convention 3

Let

$$\mathbf{y} = \psi(\mathbf{x}),\tag{23}$$

where y is an m-element vector, and x is an n-element vector. The symbol

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix}
\frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \dots & \frac{\partial y_1}{\partial x_n} \\
\frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \dots & \frac{\partial y_2}{\partial x_n} \\
\vdots & \vdots & & \vdots \\
\frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \dots & \frac{\partial y_m}{\partial x_n}
\end{bmatrix}$$
(24)

will denote the $m \times n$ matrix of first-order partial derivatives of the transformation from \mathbf{x} to \mathbf{y} . Such a matrix is called the Jacobian matrix of the transformation $\psi()$.

Notice that if \mathbf{x} is actually a scalar in Convention 3 then the resulting Jacobian matrix is a $\mathbf{m} \times 1$ matrix; that is, a single column (a vector). On the other hand, if \mathbf{y} is actually a scalar in Convention 3 then the resulting Jacobian matrix is a $1 \times \mathbf{n}$ matrix; that is, a single row (the transpose of a vector).

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{25}$$

²Much of the material in this section is extracted directly from Dhrymes (1978, Section 4.3). The interested reader is directed to this worthy reference to find additional results.

where y is $m \times 1$, x is $n \times 1$, A is $m \times n$, and A does not depend on x, then

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{A} \tag{26}$$

Proof: Since the ith element of **y** is given by

$$y_i = \sum_{k=1}^n a_{ik} x_k \tag{27}$$

it follows that

$$\frac{\partial y_i}{\partial x_j} = a_{ij} \tag{28}$$

for all $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$. Hence

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \mathbf{A} \tag{29}$$

q.e.d.

Proposition 6 Let

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{30}$$

where \mathbf{y} is $\mathbf{m} \times 1$, \mathbf{x} is $\mathbf{n} \times 1$, \mathbf{A} is $\mathbf{m} \times \mathbf{n}$, and \mathbf{A} does not depend on \mathbf{x} , as in Proposition 5. Suppose that \mathbf{x} is a function of the vector \mathbf{z} , while \mathbf{A} is independent of \mathbf{z} . Then

$$\frac{\partial \mathbf{y}}{\partial \mathbf{z}} = \mathbf{A} \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \tag{31}$$

Proof: Since the ith element of y is given by

$$y_i = \sum_{k=1}^n a_{ik} x_k \tag{32}$$

for all i = 1, 2, ..., m, it follows that

$$\frac{\partial y_i}{\partial z_j} = \sum_{k=1}^n \alpha_{ik} \frac{\partial x_k}{\partial z_j} \tag{33}$$

but the right hand side of the above is simply element (i,j) of $A_{\overline{\partial z}}^{\underline{\partial x}}$. Hence

$$\frac{\partial \mathbf{y}}{\partial \mathbf{z}} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{z}} = \mathbf{A} \frac{\partial \mathbf{x}}{\partial \mathbf{z}}$$
(34)

q.e.d.

Proposition 7 Let the scalar α be defined by

$$\alpha = \mathbf{y}^{\mathsf{T}} \mathbf{A} \mathbf{x} \tag{35}$$

where \mathbf{y} is $\mathbf{m} \times 1$, \mathbf{x} is $\mathbf{n} \times 1$, \mathbf{A} is $\mathbf{m} \times \mathbf{n}$, and \mathbf{A} is independent of \mathbf{x} and \mathbf{y} , then

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{y}^{\mathsf{T}} \mathbf{A} \tag{36}$$

and

$$\frac{\partial \alpha}{\partial \mathbf{y}} = \mathbf{x}^{\mathsf{T}} \mathbf{A}^{\mathsf{T}} \tag{37}$$

Proof: Define

$$\mathbf{w}^{\mathsf{T}} = \mathbf{y}^{\mathsf{T}} \mathbf{A} \tag{38}$$

and note that

$$\alpha = \mathbf{w}^{\mathsf{T}} \mathbf{x} \tag{39}$$

Hence, by Proposition 5 we have that

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{w}^{\mathsf{T}} = \mathbf{y}^{\mathsf{T}} \mathbf{A} \tag{40}$$

which is the first result. Since α is a scalar, we can write

$$\alpha = \alpha^{\mathsf{T}} = \mathbf{x}^{\mathsf{T}} \mathbf{A}^{\mathsf{T}} \mathbf{y} \tag{41}$$

and applying Proposition 5 as before we obtain

$$\frac{\partial \alpha}{\partial \mathbf{y}} = \mathbf{x}^{\mathsf{T}} \mathbf{A}^{\mathsf{T}} \tag{42}$$

q.e.d.

Proposition 8 For the special case in which the scalar α is given by the quadratic form

$$\alpha = \mathbf{x}^{\mathsf{T}} \mathbf{A} \mathbf{x} \tag{43}$$

where x is $n \times 1$, A is $n \times n$, and A does not depend on x, then

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{x}^{\mathsf{T}} \left(\mathbf{A} + \mathbf{A}^{\mathsf{T}} \right) \tag{44}$$

Proof: By definition

$$\alpha = \sum_{j=1}^{n} \sum_{i=1}^{n} \alpha_{ij} x_i x_j \tag{45}$$

Differentiating with respect to the kth element of \mathbf{x} we have

$$\frac{\partial \alpha}{\partial x_k} = \sum_{j=1}^n \alpha_{kj} x_j + \sum_{i=1}^n \alpha_{ik} x_i \tag{46}$$

for all k = 1, 2, ..., n, and consequently,

$$\frac{\partial \alpha}{\partial \mathbf{x}} = \mathbf{x}^{\mathsf{T}} \mathbf{A}^{\mathsf{T}} + \mathbf{x}^{\mathsf{T}} \mathbf{A} = \mathbf{x}^{\mathsf{T}} \left(\mathbf{A}^{\mathsf{T}} + \mathbf{A} \right)$$
 (47)

q.e.d.

Proposition 9 For the special case where A is a symmetric matrix and

$$\alpha = \mathbf{x}^{\mathsf{T}} \mathbf{A} \mathbf{x} \tag{48}$$

where \mathbf{x} is $\mathbf{n} \times 1$, \mathbf{A} is $\mathbf{n} \times \mathbf{n}$, and \mathbf{A} does not depend on \mathbf{x} , then

$$\frac{\partial \alpha}{\partial \mathbf{x}} = 2\mathbf{x}^{\mathsf{T}}\mathbf{A} \tag{49}$$

Proof: This is an obvious application of Proposition 8. q.e.d.

Proposition 10 Let the scalar α be defined by

$$\alpha = \mathbf{y}^{\mathsf{T}}\mathbf{x} \tag{50}$$

where y is $n \times 1$, x is $n \times 1$, and both y and x are functions of the vector z. Then

$$\frac{\partial \alpha}{\partial \mathbf{z}} = \mathbf{x}^{\mathsf{T}} \frac{\partial \mathbf{y}}{\partial \mathbf{z}} + \mathbf{y}^{\mathsf{T}} \frac{\partial \mathbf{x}}{\partial \mathbf{z}}$$
 (51)

Proof: We have

$$\alpha = \sum_{j=1}^{n} x_j y_j \tag{52}$$

Differentiating with respect to the kth element of **z** we have

$$\frac{\partial \alpha}{\partial z_k} = \sum_{j=1}^n \left(x_j \frac{\partial y_j}{\partial z_k} + y_j \frac{\partial x_j}{\partial z_k} \right) \tag{53}$$

for all k = 1, 2, ..., n, and consequently,

$$\frac{\partial \alpha}{\partial \mathbf{z}} = \frac{\partial \alpha}{\partial \mathbf{y}} \frac{\partial \mathbf{y}}{\partial \mathbf{z}} + \frac{\partial \alpha}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{z}} = \mathbf{x}^{\mathsf{T}} \frac{\partial \mathbf{y}}{\partial \mathbf{z}} + \mathbf{y}^{\mathsf{T}} \frac{\partial \mathbf{x}}{\partial \mathbf{z}}$$
(54)

q.e.d.

Proposition 11 Let the scalar α be defined by

$$\alpha = \mathbf{x}^{\mathsf{T}}\mathbf{x} \tag{55}$$

where x is $n \times 1$, and x is a function of the vector z. Then

$$\frac{\partial \alpha}{\partial \mathbf{z}} = 2\mathbf{x}^{\mathsf{T}} \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \tag{56}$$

Proof: This is an obvious application of Proposition 10. q.e.d.

Proposition 12 Let the scalar α be defined by

$$\alpha = \mathbf{y}^{\mathsf{T}} \mathbf{A} \mathbf{x} \tag{57}$$

where y is $m \times 1$, x is $n \times 1$, A is $m \times n$, and both y and x are functions of the vector z, while A does not depend on z. Then

$$\frac{\partial \alpha}{\partial \mathbf{z}} = \mathbf{x}^{\mathsf{T}} \mathbf{A}^{\mathsf{T}} \frac{\partial \mathbf{y}}{\partial \mathbf{z}} + \mathbf{y}^{\mathsf{T}} \mathbf{A} \frac{\partial \mathbf{x}}{\partial \mathbf{z}}$$
 (58)

Proof: Define

$$\mathbf{w}^{\mathsf{T}} = \mathbf{y}^{\mathsf{T}} \mathbf{A} \tag{59}$$

and note that

$$\boldsymbol{\alpha} = \mathbf{w}^{\mathsf{T}}\mathbf{x} \tag{60}$$

Applying Propositon 10 we have

$$\frac{\partial \alpha}{\partial \mathbf{z}} = \mathbf{x}^{\mathsf{T}} \frac{\partial \mathbf{w}}{\partial \mathbf{z}} + \mathbf{w}^{\mathsf{T}} \frac{\partial \mathbf{x}}{\partial \mathbf{z}}$$
 (61)

Substituting back in for \mathbf{w} we arrive at

$$\frac{\partial \alpha}{\partial \mathbf{z}} = \frac{\partial \alpha}{\partial \mathbf{y}} \frac{\partial \mathbf{y}}{\partial \mathbf{z}} + \frac{\partial \alpha}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{z}} = \mathbf{x}^{\mathsf{T}} \mathbf{A}^{\mathsf{T}} \frac{\partial \mathbf{y}}{\partial \mathbf{z}} + \mathbf{y}^{\mathsf{T}} \mathbf{A} \frac{\partial \mathbf{x}}{\partial \mathbf{z}}$$
(62)

q.e.d.

Proposition 13 Let the scalar α be defined by the quadratic form

$$\alpha = \mathbf{x}^{\mathsf{T}} \mathbf{A} \mathbf{x} \tag{63}$$

where \mathbf{x} is $\mathbf{n} \times \mathbf{1}$, \mathbf{A} is $\mathbf{n} \times \mathbf{n}$, and \mathbf{x} is a function of the vector \mathbf{z} , while \mathbf{A} does not depend on \mathbf{z} . Then

$$\frac{\partial \alpha}{\partial \mathbf{z}} = \mathbf{x}^{\mathsf{T}} \left(\mathbf{A} + \mathbf{A}^{\mathsf{T}} \right) \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \tag{64}$$

Proof: This is an obvious application of Proposition 12. q.e.d.

Proposition 14 For the special case where A is a symmetric matrix and

$$\alpha = \mathbf{x}^{\mathsf{T}} \mathbf{A} \mathbf{x} \tag{65}$$

where \mathbf{x} is $\mathbf{n} \times \mathbf{1}$, \mathbf{A} is $\mathbf{n} \times \mathbf{n}$, and \mathbf{x} is a function of the vector \mathbf{z} , while \mathbf{A} does not depend on \mathbf{z} . Then

$$\frac{\partial \alpha}{\partial \mathbf{z}} = 2\mathbf{x}^{\mathsf{T}} \mathbf{A} \frac{\partial \mathbf{x}}{\partial \mathbf{z}} \tag{66}$$

Proof: This is an obvious application of Proposition 13. q.e.d.

Definition 5 Let A be a $m \times n$ matrix whose elements are functions of the scalar parameter α . Then the derivative of the matrix A with respect to the scalar parameter α is the $m \times n$ matrix of element-by-element derivatives:

$$\frac{\partial \mathbf{A}}{\partial \alpha} = \begin{bmatrix}
\frac{\partial \mathbf{a}_{11}}{\partial \alpha} & \frac{\partial \mathbf{a}_{12}}{\partial \alpha} & \cdots & \frac{\partial \mathbf{a}_{1n}}{\partial \alpha} \\
\frac{\partial \mathbf{a}_{21}}{\partial \alpha} & \frac{\partial \mathbf{a}_{22}}{\partial \alpha} & \cdots & \frac{\partial \mathbf{a}_{2n}}{\partial \alpha} \\
\vdots & \vdots & & \vdots \\
\frac{\partial \mathbf{a}_{m1}}{\partial \alpha} & \frac{\partial \mathbf{a}_{m2}}{\partial \alpha} & \cdots & \frac{\partial \mathbf{a}_{mn}}{\partial \alpha}
\end{bmatrix}$$
(67)

Proposition 15 Let A be a nonsingular, $m \times m$ matrix whose elements are functions of the scalar parameter α . Then

$$\frac{\partial \mathbf{A}^{-1}}{\partial \alpha} = -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \alpha} \mathbf{A}^{-1} \tag{68}$$

Proof: Start with the definition of the inverse

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I} \tag{69}$$

and differentiate, yielding

$$\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \alpha} + \frac{\partial \mathbf{A}^{-1}}{\partial \alpha} \mathbf{A} = \mathbf{0}$$
 (70)

rearranging the terms yields

$$\frac{\partial \mathbf{A}^{-1}}{\partial \alpha} = -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \alpha} \mathbf{A}^{-1} \tag{71}$$

q.e.d.

6 References

- Dhrymes, Phoebus J., 1978, **Mathematics for Econometrics**, Springer-Verlag, New York, 136 pp.
- Golub, Gene H., and Charles F. Van Loan, 1983, **Matrix Computations**, Johns Hopkins University Press, Baltimore, Maryland, 476 pp.
- Graybill, Franklin A., 1983, Matrices with Applications in Statistics, 2nd Edition, Wadsworth International Group, Belmont, California, 461 pp.