Matrix Differentiation

(and some other stuff)

Randal J. Barnes
Department of Civil Engineering, University of Minnesota
Minneapolis, 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 $a_{ij} \in \Re$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$. Then the ordered rectangular array

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

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,

$$\mathbf{A} = [a_{ij}], \qquad i = 1, 2, \ldots, m; \; j = 1, 2, \ldots, n, \tag{2}$$

denotes a matrix with m rows and n columns, whose typical element is a_{ij}. Note, the first subscript locates the *row* in which the typical element lies while the second subscript locates the *column*. For example, 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, $\underset{m\times n}{\mathbf{A}}$, indicates a known, multi-column matrix with m rows and n columns.

A superscript $^{\mathsf{T}}$ denotes the matrix transpose operation; for example, $\mathbf{A}^{\mathsf{T}}$ 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 $\mathrm{rank}(\mathbf{A})$. An identity matrix will be denoted by $\mathbf{I}$, and $\mathbf{0}$ will denote a null matrix.

3 Matrix Multiplication

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

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

then $\mathbf{C}$ 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 $i = 1, 2, \ldots, m, \quad j = 1, 2, \ldots, p$.

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

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

is given by

$$z_i = \sum_{k=1}^{n} a_{ik} x_k \tag{6}$$

for all $i = 1, 2, \ldots, m$. Similarly, let $\mathbf{y}$ 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

$$z_i = \sum_{k=1}^{n} a_{ki} y_k \tag{8}$$

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

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

is given by

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

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

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

$$\mathbf{C} = \mathbf{AB} \tag{11}$$

then

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

Proof*: The typical element of* $\mathbf{C}$ *is given by*

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

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

$$d_{ij} = c_{ji} = \sum_{k=1}^{n} a_{jk}b_{ki} \tag{14}$$

Hence,

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

q.e.d.

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

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

then

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

Proof*:*

$$\mathbf{CB}^{-1}\mathbf{A}^{-1} = \mathbf{ABB}^{-1}\mathbf{A}^{-1} = \mathbf{I} \tag{18}$$

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 $\mathbf{A}$ be $m \times n$ and write

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

where $\mathbf{B}$ is $m_1 \times n_1$, $\mathbf{E}$ is $m_2 \times n_2$, $\mathbf{C}$ is $m_1 \times n_2$, $\mathbf{D}$ 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 $\mathbf{A}$.

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

Proposition 4 *Let* $\mathbf{A}$ *be a square, nonsingular matrix of order* m*. Partition* $\mathbf{A}$ *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* m_1*,* $\mathbf{A}_{22}$ *is a nonsingular matrix of order* m_2*, and* $m_1 + m_2 = m$*. Then*

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

Proof*: Direct multiplication of the proposed* $\mathbf{A}^{-1}$ *and* $\mathbf{A}$ *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 $\mathbf{y}$ is an m-element vector, and $\mathbf{x}$ is an n-element vector. The symbol

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

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 $m \times 1$ 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 $1 \times n$ matrix; that is, a single row (the transpose of a vector).

Proposition 5 *Let*

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

[2]Much 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 $\mathbf{y}$ *is* $m \times 1$, $\mathbf{x}$ *is* $n \times 1$, $\mathbf{A}$ *is* $m \times n$, *and* $\mathbf{A}$ *does not depend on* $\mathbf{x}$, *then*

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

Proof*: Since the* i*th element of* $\mathbf{y}$ *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 $i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n$. *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* $m \times 1$, $\mathbf{x}$ *is* $n \times 1$, $\mathbf{A}$ *is* $m \times n$, *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* i*th element of* $\mathbf{y}$ *is given by*

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

for all $i = 1, 2, \dots, m$, *it follows that*

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

but the right hand side of the above is simply element (i, j) *of* $\mathbf{A}\frac{\partial \mathbf{x}}{\partial \mathbf{z}}$. *Hence*

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

q.e.d.

Proposition 7 *Let the scalar* α *be defined by*

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

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

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

and

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

Proof*: Define*

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

and note that

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

Hence, by Proposition 5 we have that

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

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

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

and applying Proposition 5 as before we obtain

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

q.e.d.

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

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

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

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

Proof*: By definition*

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

Differentiating with respect to the k*th element of* $\mathbf{x}$ *we have*

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

for all $k = 1, 2, \ldots, n$, *and consequently,*

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

q.e.d.

$f(x) = \frac{1}{2} x^T A x + b^T x + c$

Proposition 9 *For the special case where* $\mathbf{A}$ *is a symmetric matrix and*

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

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

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

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

Proposition 10 *Let the scalar* α *be defined by*

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

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

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

Proof*: We have*

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

Differentiating with respect to the k*th element of* $\mathbf{z}$ *we have* 元素

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

for all $k = 1, 2, \ldots, n$, *and consequently,*

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

q.e.d.

Proposition 11 *Let the scalar* α *be defined by*

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

二乘

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

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

$L(w) = (y - Xw)^T (y - Xw)$

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

Proposition 12 *Let the scalar* α *be defined by*

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

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

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

Proof*: Define*

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

and note that

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

Applying Propositon 10 we have

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

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

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

q.e.d.

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

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

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

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

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

Proposition 14 *For the special case where* $\mathbf{A}$ *is a symmetric matrix and*

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

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

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

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

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

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

Proposition 15 *Let* $\mathbf{A}$ *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

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

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.