Maybeck (1979)

13. Maybeck, P. S., "Applied Optimal Estimation—Kalman Filter Design and Implementation," notes for a continuing education course offered by the Air Force Institute of Technology, Wright-Patterson AFB, Ohio, semiannually since December 1974.
14. Meditch, J. S., *Stochastic Optimal Linear Estimation and Control*. McGraw-Hill, New York, 1969.
15. McGarty, T. P., *Stochastic Systems and State Estimation*. Wiley, New York, 1974.
16. Sage, A. P., and Melsa, J. L., *Estimation Theory with Application to Communications and Control*. McGraw-Hill, New York, 1971.
17. Schweppe, F. C., *Uncertain Dynamic Systems*. Prentice-Hall, Englewood Cliffs, New Jersey, 1973.
18. Van Trees, H. L., *Detection, Estimation and Modulation Theory*, Vol. 1. Wiley, New York, 1968.

APPENDIX AND PROBLEMS

Matrix Analysis

This appendix and its associated problems present certain results from elementary matrix analysis, as well as notation conventions, that will be of use throughout the text. If the reader desires more than this brief review, the list of references [1–11] at the end provides a partial list of good sources.

A.1 Matrices

An *n-by-m matrix* is a rectangular array of scalars consisting of n rows and m columns, denoted by a boldfaced capitalized letter, as

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1m} \\ A_{21} & A_{22} & \cdots & A_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nm} \end{bmatrix}$$

Thus, A_{ij} is the scalar element in the ith row and jth column of **A**, and unless specified otherwise, will be assumed herein to be a real number (or a real-valued scalar function).

If all of the elements A_{ij} are zeros, **A** is called a *zero matrix* or *null matrix*, denoted as **0**.

If all of the elements of an n-by-n (*square*) matrix are zeros except for those along the principal diagonal, as

$$\mathbf{A} = \begin{bmatrix} A_{11} & 0 & \cdots & 0 \\ 0 & A_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{nn} \end{bmatrix}$$

the **A** is called *diagonal*. Furthermore, if $A_{ii} = 1$ for all i, the matrix is called the *identity matrix* and is denoted by **I**.

A square matrix is *symmetric* if $A_{ij} = A_{ji}$ for all values of i and j from 1 to n. Thus, a diagonal matrix is always symmetric. Show that there are at most $\frac{1}{2}n(n+1)$ nonredundant elements in an n-by-n symmetric matrix.

A *lower triangular* matrix is a square matrix, all of whose elements above the principal diagonal are zero, as

$$\mathbf{A} = \begin{bmatrix} A_{11} & 0 & \cdots & 0 \\ A_{21} & A_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix}$$

Similarly, an *upper triangular* matrix is a square matrix with all zeros below the principal diagonal.

A matrix composed of a single column, i.e., an n-by-1 matrix, is called an n-dimensional *vector* or *n-vector* and will be denoted by a boldfaced lower case letter, as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Thus, x_i is the ith scalar element, or "component," of the n-vector **x**. (The directed line segment from the origin to a point in Euclidean n-dimensional space can be represented, relative to a chosen basis or reference coordinate directions, by **x**, and then x_i is the component along the ith basis vector or reference direction.) Properties of general nonsquare matrices (as described in Sections A.2, A.3, and A.10 to follow) are true specifically for vectors.

A matrix can be subdivided not only into its scalar elements, but also into arrays of elements called matrix *partitions*, such as

$$\mathbf{A} = \begin{array}{c} \left[\begin{array}{c|c|c|c} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1i} \\ \hline \vdots & \vdots & & \vdots \\ \hline \mathbf{A}_{k1} & \mathbf{A}_{k2} & \cdots & \mathbf{A}_{ki} \end{array}\right] \begin{array}{l} \}\, n_1 \text{ rows} \\ \\ \}\, n_k \text{ rows} \end{array} \\ \begin{array}{cccc} \underbrace{\quad}_{m_1 \text{ columns}} & \underbrace{\quad}_{m_2 \text{ columns}} & & \underbrace{\quad}_{m_i \text{ columns}} \end{array} \end{array} \qquad \mathbf{x} = \left[\begin{array}{c} \mathbf{x}_1 \\ \hline \mathbf{x}_2 \\ \hline \vdots \\ \hline \mathbf{x}_k \end{array}\right] \begin{array}{l} \}\, n_1 \text{ components} \\ \}\, n_2 \text{ components} \\ \\ \}\, n_k \text{ components} \end{array}$$

A square matrix **A** is termed *block diagonal* if it can be subdivided into partitions such that $\mathbf{A}_{ij} = \mathbf{0}$ for all partitions for which $i \neq j$, and such that all partitions $\mathbf{A}_{ii}$ are square.

A.2 Equality, Addition, and Multiplication

Two n-by-m matrices **A** and **B** are *equal* if and only if $A_{ij} = B_{ij}$ for all i and j.

If **A** and **B** are both n-by-m matrices, their *sum* can be defined as $\mathbf{C} = \mathbf{A} + \mathbf{B}$, where **C** is an n-by-m matrix whose elements satisfy $C_{ij} = A_{ij} + B_{ij}$ for all i and j.

Their *difference* would be defined similarly. Show that

(a) $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$.
(b) $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$.
(c) $\mathbf{A} + \mathbf{0} = \mathbf{0} + \mathbf{A} = \mathbf{A}$.

The product of an n-by-m matrix $\mathbf{A}$ by a scalar b is the n-by-m matrix $\mathbf{C} = b\mathbf{A} = \mathbf{A}b$ composed of elements $C_{ij} = bA_{ij}$.

If $\mathbf{A}$ is n-by-m and $\mathbf{B}$ is m-by-r, then the *product* $\mathbf{C} = \mathbf{AB}$ can be defined as an n-by-r matrix with elements $C_{ij} = \sum_{k=1}^{m} A_{ik}B_{kj}$. This product can be defined only if the number of columns of $\mathbf{A}$ equals the number of rows of $\mathbf{B}$: only if $\mathbf{A}$ and $\mathbf{B}$ are "conformable" for the product $\mathbf{AB}$. Thus, the ordering in the product is important, and $\mathbf{AB}$ can be described as "premultiplying" $\mathbf{B}$ by $\mathbf{A}$ or "postmultiplying" $\mathbf{A}$ by $\mathbf{B}$. Show that for general conformable matrices

(d) $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$.
(e) $\mathbf{IA} = \mathbf{AI} = \mathbf{A}$.
(f) $\mathbf{0A} = \mathbf{A0} = \mathbf{0}$.
(g) $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$.
(h) in general, $\mathbf{AB} \neq \mathbf{BA}$, even for $\mathbf{A}$ and $\mathbf{B}$ both square.
(i) $\mathbf{AB} = \mathbf{0}$ in general does not imply that $\mathbf{A}$ or $\mathbf{B}$ is $\mathbf{0}$.

A product of particular importance is that of an n-by-m matrix $\mathbf{A}$ with an m-vector $\mathbf{x}$ to yield an n-vector $\mathbf{y} = \mathbf{Ax}$, with components $y_i = \sum_{j=1}^{m} A_{ij}x_j$. Such a matrix multiplication can be used to represent a linear transformation of a vector. More general functions, not expressible through matrix multiplications, can be written as

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) \leftrightarrow \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} f_1(x_1, x_2, \ldots, x_m) \\ \vdots \\ f_n(x_1, x_2, \ldots, x_m) \end{bmatrix}$$

Matrix operations upon partitioned matrices obey the same rules of equality, addition, and multiplication, provided that the matrix partitions are conformable. For instance, show that

(j)

$$\left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right] + \left[\begin{array}{c|c} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \hline \mathbf{B}_{21} & \mathbf{B}_{22} \end{array}\right] = \left[\begin{array}{c|c} \mathbf{A}_{11} + \mathbf{B}_{11} & \mathbf{A}_{12} + \mathbf{B}_{12} \\ \hline \mathbf{A}_{21} + \mathbf{B}_{21} & \mathbf{A}_{22} + \mathbf{B}_{22} \end{array}\right]$$

(k)

$$\left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right] \cdot \left[\begin{array}{c|c} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \hline \mathbf{B}_{21} & \mathbf{B}_{22} \end{array}\right] = \left[\begin{array}{c|c} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \hline \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{array}\right]$$

A.3 Transposition

The *transpose* of an n-by-m matrix $\mathbf{A}$ is the m-by-n matrix denoted as $\mathbf{A}^{\mathrm{T}}$ that satisfies $A^{\mathrm{T}}_{ij} = A_{ji}$ for all i and j. Thus, transposition can be interpreted as interchanging the roles of rows and columns of a matrix. For example, if $\mathbf{x}$ is an n-vector, $\mathbf{x}^{\mathrm{T}}$ is a 1-by-n matrix, or "row vector." Show that

(a) $(\mathbf{A}^{\mathrm{T}})^{\mathrm{T}} = \mathbf{A}$.
(b) $(\mathbf{A} + \mathbf{B})^{\mathrm{T}} = \mathbf{A}^{\mathrm{T}} + \mathbf{B}^{\mathrm{T}}$.
(c) $(\mathbf{AB})^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$.
(d) if $\mathbf{A}$ is a symmetric matrix, $\mathbf{A}^{\mathrm{T}} = \mathbf{A}$.
(e) if $\mathbf{x}$ and $\mathbf{y}$ are n-vectors, $\mathbf{x}^{\mathrm{T}}\mathbf{y}$ is a scalar and $\mathbf{xy}^{\mathrm{T}}$ is a square n-by-n matrix; $\mathbf{xx}^{\mathrm{T}}$ is symmetric as well.
(f) if $\mathbf{A}$ is a symmetric n-by-n matrix and $\mathbf{B}$ is a general m-by-n matrix, then $\mathbf{C} = \mathbf{BAB}^{\mathrm{T}}$ is a symmetric m-by-m matrix.
(g) if $\mathbf{A}$ and $\mathbf{B}$ are both symmetric n-by-n matrices, $(\mathbf{A} + \mathbf{B})$ is also symmetric but $(\mathbf{AB})$ generally is not.
(h)

$$\left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array}\right]^{\mathrm{T}} = \left[\begin{array}{c|c} \mathbf{A}^{\mathrm{T}} & \mathbf{C}^{\mathrm{T}} \\ \hline \mathbf{B}^{\mathrm{T}} & \mathbf{D}^{\mathrm{T}} \end{array}\right]$$

A.4 Matrix Inversion, Singularity, and Determinants

Given a square matrix $\mathbf{A}$, if there exists a matrix such that both premultiplying and postmultiplying it by $\mathbf{A}$ yields the identity, then this matrix is called the *inverse* of $\mathbf{A}$, and is denoted by $\mathbf{A}^{-1}$: $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. A square matrix that does not possess such an inverse is said to be *singular*. If $\mathbf{A}$ has an inverse, the inverse is unique, and $\mathbf{A}$ is termed *nonsingular*. Show that

(a) if $\mathbf{A}$ is nonsingular, then so is $\mathbf{A}^{-1}$, and $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$.
(b) $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ if all indicated inverses exist.
(c) $(\mathbf{A}^{-1})^{\mathrm{T}} = (\mathbf{A}^{\mathrm{T}})^{-1}$.
(d) if a transformation of variables is represented by $\mathbf{x}^* = \mathbf{Ax}$ and if $\mathbf{A}^{-1}$ exists, then $\mathbf{x} = \mathbf{A}^{-1}\mathbf{x}^*$.

The *determinant* of a square n-by-n matrix $\mathbf{A}$ is a scalar-valued function of the matrix elements, denoted by $|\mathbf{A}|$, the evaluation of which can be performed recursively through $|\mathbf{A}| = \sum_{j=1}^{n} A_{ij}C_{ij}$ for any fixed $i = 1, 2, \ldots,$ or n; C_{ij} is the "cofactor" of A_{ij}, defined through $C_{ij} = (-1)^{i+j}M_{ij}$, and M_{ij} is the "minor" of A_{ij}, defined as the determinant of the $(n-1)$-by-$(n-1)$ matrix formed by deleting the ith row and jth column of the n-by-n $\mathbf{A}$. (Note that iterative application of these relationships ends with the evaluation of determinants of 1-by-1 matrices or scalars as the scalar values themselves.) Show that

(e) if $\mathbf{A}$ is 2-by-2, then $|\mathbf{A}| = A_{11}A_{22} - A_{12}A_{21}$.

(f) if $\mathbf{A}$ is 3-by-3, then

$$|\mathbf{A}| = A_{11}A_{22}A_{33} + A_{12}A_{23}A_{31} + A_{13}A_{32}A_{21} - A_{11}A_{32}A_{23} - A_{12}A_{21}A_{33} - A_{13}A_{22}A_{31}$$

(g) $|\mathbf{A}^T| = |\mathbf{A}|$.
(h) if all the elements of any row or column of $\mathbf{A}$ are zero, $|\mathbf{A}| = 0$.
(i) if any row (column) of $\mathbf{A}$ is a multiple of any other row (column), then $|\mathbf{A}| = 0$.
(j) if a scalar multiple of any row (column) is added to any other row (column) of $\mathbf{A}$, the value of the determinant is unchanged.
(k) if $\mathbf{A}$ and $\mathbf{B}$ are n-by-n, $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$.
(l) if $\mathbf{A}$ is diagonal, then $|\mathbf{A}|$ equals the product of its diagonal elements: $|\mathbf{A}| = \prod_{i=1}^{n} A_{ii}$.
(m) if the n-by-n $\mathbf{A}$ is nonsingular, then $|\mathbf{A}| \neq 0$ and $\mathbf{A}^{-1}$ can be evaluated as $\mathbf{A}^{-1} = [\operatorname{adj}\mathbf{A}]/|\mathbf{A}|$, where $[\operatorname{adj}\mathbf{A}]$ is the adjoint of $\mathbf{A}$, defined as the n-by-n matrix whose ij element (i.e., in the ith row and jth column) is the cofactor C_{ji}.
(n) $|\mathbf{A}^{-1}| = 1/|\mathbf{A}|$ if $|\mathbf{A}| \neq 0$.
(o) $|\mathbf{A}| = 0$ if and only if $\mathbf{A}$ is singular.
(p)

$$\left|\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{C} \end{bmatrix}\right| = |\mathbf{A}||\mathbf{C}|$$

If $\mathbf{A}$ is such that its inverse equals its transpose, $\mathbf{A}^{-1} = \mathbf{A}^T$, then $\mathbf{A}$ is termed *orthogonal*. If $\mathbf{A}$ is orthogonal, $\mathbf{AA}^T = \mathbf{A}^T\mathbf{A} = \mathbf{I}$, and $|\mathbf{A}| = \pm 1$.

A.5 *Linear Independence and Rank*

A set of k n-vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ is said to be *linearly dependent* if there exists a set of k constants $c_1, c_2, \ldots, c_k$ (at least one of which is not zero) such that $\sum_{i=1}^{k} c_i\mathbf{x}_i = \mathbf{0}$. If no such set of constants exists, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ are said to be *linearly independent*.

The *rank* of an n-by-m matrix is the order of the largest square nonsingular matrix that can be formed by deleting rows and columns. Show that

(a) if the m-by-n $\mathbf{A}$ is partitioned into column vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ and $\mathbf{x}$ is an n-vector, then $\mathbf{Ax} = \sum_{i=1}^{n} \mathbf{a}_i x_i$.
(b) the rank of $\mathbf{A}$ equals the number of linearly independent rows or columns of $\mathbf{A}$, whichever is smaller.
(c) if $\mathbf{A}$ is n-by-n, then it is of rank n (of "full rank") if and only if it is nonsingular.
(d) the rank of $\mathbf{xx}^T$ is one.

A.6 *Eigenvalues and Eigenvectors*

The equation $\mathbf{Ax} = \lambda\mathbf{x}$ for n-by-n $\mathbf{A}$, or $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$, possesses a nontrivial solution if an only if $|\mathbf{A} - \lambda\mathbf{I}| = 0$. The nth order polynomial $f(\lambda) = |\mathbf{A} - \lambda\mathbf{I}|$ is called the *characteristic polynomial* of $\mathbf{A}$, and the equation $f(\lambda) = 0$ is called its *characteristic equation*. The n *eigenvalues* of $\mathbf{A}$ are the (not necessarily distinct) roots of this equation, and the nonzero solutions to $\mathbf{Ax}_i = \lambda_i\mathbf{x}_i$, corresponding to the roots λ_i, are called *eigenvectors*. It can be shown that $|\mathbf{A}|$ equals the product of the eigenvalues of $\mathbf{A}$, and $\sum_{i=1}^{n} A_{ii} = \sum_{i=1}^{n} \lambda_i$.

Let the eigenvalues of the n-by-n $\mathbf{A}$ be the distinct values $\lambda_1, \lambda_2, \ldots, \lambda_n$, and let the associated eigenvectors be $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$. Then, if $\mathbf{E} = [\mathbf{e}_1 | \mathbf{e}_2 | \cdots | \mathbf{e}_n]$, $\mathbf{E}$ is nonsingular, and $\mathbf{E}^{-1}\mathbf{AE}$ is a diagonal matrix whose ith diagonal element is λ_i, for $i = 1, 2, \ldots, n$. Moreover, if $\mathbf{A}$ is also symmetric, then the eigenvalues are all real and $\mathbf{E}$ is orthogonal.

(a) Obtain the eigenvalues and eigenvectors for a general 2-by-2 $\mathbf{A}$; generate $\mathbf{E}$ and $\mathbf{E}^{-1}\mathbf{AE}$.
(b) Repeat for a general symmetric 2-by-2 $\mathbf{A}$; show that λ_1 and λ_2 must be real, and that $\mathbf{E}$ is orthogonal.
(c) Show that $|\mathbf{A}| = 0$ if and only if at least one eigenvalue is zero.

A.7 *Quadratic Forms and Positive (Semi-) Definiteness*

If $\mathbf{A}$ is n-by-n and $\mathbf{x}$ is an n-vector, then the scalar quantity $\mathbf{x}^T\mathbf{Ax}$ is called a *quadratic form*. Show that

(a) $\mathbf{x}^T\mathbf{Ax} = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij}x_ix_j$.
(b) without loss of generality, $\mathbf{A}$ can always be considered to be symmetric, since if $\mathbf{A}$ is not symmetric, a symmetric matrix $\mathbf{B}$ can always be defined by

$$B_{ij} = \begin{cases} A_{ij} & i = j \\ \frac{1}{2}(A_{ij} + A_{ji}) & i \neq j \end{cases}$$

for i and j equal to $1, 2, \ldots, n$, and then $\mathbf{x}^T\mathbf{Ax} = \mathbf{x}^T\mathbf{Bx}$.
(c) if $\mathbf{A}$ is diagonal, $\mathbf{x}^T\mathbf{Ax} = \sum_{i=1}^{n} A_{ii}x_i^2$.

If $\mathbf{x}^T\mathbf{Ax} > 0$ for all $\mathbf{x} \neq 0$, the quadratic form is said to be *positive definite*, as is the matrix $\mathbf{A}$ itself, often written notationally as $\mathbf{A} > \mathbf{0}$. If $\mathbf{x}^T\mathbf{Ax} \geq 0$ for all $\mathbf{x} \neq \mathbf{0}$, the quadratic form and matrix $\mathbf{A}$ are termed *positive semidefinite*, denoted as $\mathbf{A} \geq \mathbf{0}$. Furthermore, the notation $\mathbf{A} > \mathbf{B}$ ($\mathbf{A} \geq \mathbf{B}$) is meant to say that $(\mathbf{A} - \mathbf{B})$ is positive definite (semidefinite). Show that

(d) if $\mathbf{A}$ is positive definite, it is nonsingular, and its inverse $\mathbf{A}^{-1}$ is also positive definite.

(e) the symmetric $\mathbf{A}$ is positive definite if and only if its eigenvalues are all positive; $\mathbf{A}$ is positive semidefinite if and only if its eigenvalues are all positive or zero.

(f) if $\mathbf{A}$ is positive definite and $\mathbf{B}$ is positive semidefinite, $(\mathbf{A}+\mathbf{B})$ is positive definite.

A.8 Trace

The *trace* of an n-by-n matrix $\mathbf{A}$, denoted as $\mathrm{tr}(\mathbf{A})$, is defined as the sum of the diagonal terms:

$$\mathrm{tr}(\mathbf{A}) \triangleq \sum_{i=1}^{n} A_{ii}$$

Using this basic definition, show that

(a) $\mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{A}^{\mathrm{T}})$.
(b) $\mathrm{tr}(\mathbf{A}_1 + \mathbf{A}_2) = \mathrm{tr}(\mathbf{A}_1) + \mathrm{tr}(\mathbf{A}_2)$.
(c) if $\mathbf{B}$ is n-by-m and $\mathbf{C}$ is m-by-n, so that $\mathbf{BC}$ is n-by-n and $\mathbf{CB}$ is m-by-m, then

$$\mathrm{tr}(\mathbf{BC}) = \mathrm{tr}(\mathbf{CB}) = \mathrm{tr}(\mathbf{B}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}) = \mathrm{tr}(\mathbf{C}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}})$$

(d) if $\mathbf{x}$ and $\mathbf{y}$ are n-vectors and $\mathbf{A}$ is n-by-n, then

$$\mathrm{tr}(\mathbf{x}\mathbf{y}^{\mathrm{T}}) = \mathrm{tr}(\mathbf{x}^{\mathrm{T}}\mathbf{y}) = \mathbf{x}^{\mathrm{T}}\mathbf{y}$$
$$\mathrm{tr}(\mathbf{A}\mathbf{x}\mathbf{y}^{\mathrm{T}}) = \mathrm{tr}(\mathbf{y}^{\mathrm{T}}\mathbf{A}\mathbf{x}) = \mathbf{y}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{y}$$

A.9 Similarity

If $\mathbf{A}$ and $\mathbf{B}$ are n-by-n and $\mathbf{T}$ is a nonsingular n-by-n matrix, and $\mathbf{A} = \mathbf{T}^{-1}\mathbf{B}\mathbf{T}$, then $\mathbf{A}$ and $\mathbf{B}$ are said to be related by a *similarity transformation*, or are simply termed *similar*. Show that

(a) if $\mathbf{A} = \mathbf{T}^{-1}\mathbf{B}\mathbf{T}$, then $\mathbf{B} = \mathbf{T}\mathbf{A}\mathbf{T}^{-1}$.
(b) if $\mathbf{A}$ and $\mathbf{B}$ are similar, their determinants, eigenvalues, eigenvectors, characteristic polynomials, and traces are equal; also if $\mathbf{A}$ is positive definite, so is $\mathbf{B}$ and vice versa.

A.10 Differentiation and Integration

Let $\mathbf{A}$ be an n-by-m matrix function of a scalar variable t, such as time. Then $d\mathbf{A}/dt \triangleq \dot{\mathbf{A}}(t)$ is defined as the n-by-m matrix with ij element as dA_{ij}/dt for all i and j; $\int \mathbf{A}(\tau)\,d\tau$ is defined similarly as a matrix composed of elements $\int A_{ij}(\tau)\,d\tau$. Derivatives and integrals of vectors are special cases of these definitions. Show that

(a) $d[\mathbf{A}^{\mathrm{T}}(t)]/dt = [d\mathbf{A}(t)/dt]^{\mathrm{T}}$ and similarly for integration.
(b) $d[\mathbf{A}(t)\mathbf{B}(t)]/dt = \dot{\mathbf{A}}(t)\mathbf{B}(t) + \mathbf{A}(t)\dot{\mathbf{B}}(t)$.

Let the scalar s and the n-vector $\mathbf{x}$ be functions of the m-vector $\mathbf{v}$. By convention, the following derivative definitions are made:

$$\frac{\partial s}{\partial \mathbf{v}} = \begin{bmatrix} \dfrac{\partial s}{\partial v_1} & \dfrac{\partial s}{\partial v_2} & \cdots & \dfrac{\partial s}{\partial v_m} \end{bmatrix}$$

$$\frac{\partial \mathbf{x}}{\partial \mathbf{v}} = \left[\frac{\partial \mathbf{x}}{\partial v_1} \;\Big|\; \frac{\partial \mathbf{x}}{\partial v_2} \;\Big|\; \cdots \;\Big|\; \frac{\partial \mathbf{x}}{\partial v_m} \right] = \begin{bmatrix} \dfrac{\partial x_1}{\partial v_1} & \dfrac{\partial x_1}{\partial v_2} & \cdots & \dfrac{\partial x_1}{\partial v_m} \\ \dfrac{\partial x_2}{\partial v_1} & \dfrac{\partial x_2}{\partial v_2} & \cdots & \dfrac{\partial x_2}{\partial v_m} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial x_n}{\partial v_1} & \dfrac{\partial x_n}{\partial v_2} & \cdots & \dfrac{\partial x_n}{\partial v_m} \end{bmatrix}$$

By generating the appropriate forms for scalar components and recombining, show the validity of the following useful forms (for the vectors $\mathbf{x}$ and $\mathbf{y}$ assumed to be functions of $\mathbf{v}$ possibly, and the vector $\mathbf{z}$ and matrices $\mathbf{A}$ and $\mathbf{B}$ assumed constant):

(c) $\partial\mathbf{v}/\partial\mathbf{v} = \mathbf{I}$.
(d) $\partial(\mathbf{A}\mathbf{x})/\partial\mathbf{v} = \mathbf{A}\,\partial\mathbf{x}/\partial\mathbf{v}$, and thus, $\partial(\mathbf{A}\mathbf{v})/\partial\mathbf{v} = \mathbf{A}$.
(e) $\partial(\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{y})/\partial\mathbf{v} = \mathbf{x}^{\mathrm{T}}\mathbf{A}\,\partial\mathbf{y}/\partial\mathbf{v} + \mathbf{y}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\,\partial\mathbf{x}/\partial\mathbf{v}$

and so

$$\partial(\mathbf{z}^{\mathrm{T}}\mathbf{A}\mathbf{v})/\partial\mathbf{v} = \mathbf{z}^{\mathrm{T}}\mathbf{A}, \qquad \partial(\mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{z})/\partial\mathbf{v} = \mathbf{z}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$$

and

$$\partial(\mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{v})/\partial\mathbf{v} = \mathbf{v}^{\mathrm{T}}\mathbf{A} + \mathbf{v}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}} = 2\mathbf{v}^{\mathrm{T}}\mathbf{A} \qquad \text{if} \quad \mathbf{A} = \mathbf{A}^{\mathrm{T}}$$

and

$$\partial\{(\mathbf{z}-\mathbf{B}\mathbf{v})^{\mathrm{T}}\mathbf{A}(\mathbf{z}-\mathbf{B}\mathbf{v})\}/\partial\mathbf{v} = -2(\mathbf{z}-\mathbf{B}\mathbf{v})^{\mathrm{T}}\mathbf{A}\mathbf{B} \qquad \text{if} \quad \mathbf{A} = \mathbf{A}^{\mathrm{T}}.$$

REFERENCES

1. Bellman, R. E., *Introduction to Matrix Analysis*. McGraw-Hill, New York, 1960.
2. DeRusso, P. M., Roy, R. J., and Close, C. M., *State Variables for Engineers*. Wiley, New York, 1965.
3. Edelen, D. G. B., and Kydoniefs, A. D., *An Introduction to Linear Algebra for Science and Engineering* (2nd Ed.). American Elsevier, New York, 1976.