Technical Note 1965-53

M. Athans
F. C. Schweppe

Gradient Matrices and Matrix Calculations

17 November 1965

Prepared under Electronic Systems Division Contract AF 19(628)-5167 by

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Lexington, Massachusetts

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

GRADIENT MATRICES AND MATRIX CALCULATIONS

M. ATHANS

Consultant

F. C. SCHWEPPE

Group 22

TECHNICAL NOTE 1965-53

17 NOVEMBER 1965

LEXINGTON MASSACHUSETTS

ABSTRACT

The purpose of this report is to define a useful shorthand notation for dealing with matrix functions and to use these results in order to compute the gradient matrices of several scalar functions of matrices.

Accepted for the Air Force
Stanley J. Wisniewski
Lt Colonel, USAF
Chief, Lincoln Laboratory Office

TABLE OF CONTENTS

1. INTRODUCTION

The purpose of this report is to present a shorthand notation for matrix manipulation and formulae of differentiation for matrix quantities. The shorthand and formulae are especially useful whenever one deals with the analysis and control of dynamical systems which are described by matrix differential equations. There are other areas of application but the control of matrix differential equations provided the motivation for this study. References [1] through [6] deal with the analysis and control of dynamical systems which are described by matrix differential equations.

Much of the material presented in this report is available elsewhere in different forms; it is summarized herein for the sake of convenience. Two references were used extensively for the mathematical background; these are Bodewig (Reference [7]) and Bellman, (Reference [8]).

The organization of the report is as follows: In Section 2 we present the definitions of the unit vectors $\underline{e}_i$ and of the unit matrices $\underline{E}_{ij}$. In Section 3 we indicate the use of the matrices $\underline{E}_{ij}$ as basis in the space of $n \times n$ matrices. In Section 4 we present several relations which can be used to decompose a given matrix into its column and row vectors. Section 5 deals with operations involving the unit vectors $\underline{e}_i$ and the unit matrices $\underline{E}_{ij}$. In Section 6 we show how the trace function can be used to represent the scalar product of two matrices. In Section 7 we define the differentials of a vector and of a matrix and we also define the motion of a gradient matrix. Section 8 contains a variety of formulae for the gradient matrix of trace functions. Section 9 contains relations for the gradient matrix of determinant functions. Section 10 contains relations involving partitioned matrices. A table summarizing the gradient formulae of Sections 8 and 9 is also provided.

2. NOTATION

Throughout this report column vectors will be denoted by underlined letters and matrices by underlined capital letters. The prime (′) will denote transposition.

A column vector $\underline{v}$ with components $v_1, v_2, \ldots, v_n$ is

$$\underline{v} = \begin{bmatrix} v_1 \\ v_2 \\ \cdot \\ \cdot \\ \cdot \\ v_n \end{bmatrix} . \tag{2.1}$$

In particular, the unit vectors $\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_n$ are defined as follows:

$$\underline{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix} , \quad \underline{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} , \ldots, \quad \underline{e}_n = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} . \tag{2.2}$$

An $n \times m$ matrix $\underline{A}$ with elements a_{ij} $(i = 1, 2, \ldots, n \; ; \; j = 1, 2, \ldots, m)$ is denoted by

$$\underline{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} . \tag{2.3}$$

If $m = n$, then $\underline{A}$ is square. If $\underline{A} = \underline{A}'$ then $\underline{A}$ is symmetric.

The unit matrices $\underline{E}_{ij}$ are square matrices such that all their elements are zero, except the one located at the i-th row and j-th column which is unity. For example,

$$\underline{E}_{12} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & & \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} \quad . \tag{2.4}$$

The unit matrix $\underline{E}_{ij}$ is related to the unit vectors $\underline{e}_i$ and $\underline{e}_j$ as follows:

$$\underline{E}_{ij} = \underline{e}_i \, \underline{e}_j' \quad . \tag{2.5}$$

The identity matrix $\underline{I}$,

$$\underline{I} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ 0 & 0 & \dots & 1 \end{bmatrix} \tag{2.6}$$

can thus be written

$$\underline{I} = \sum_{i=1}^{n} \underline{E}_{ii} = \sum_{i=1}^{n} \underline{e}_i \, \underline{e}_i' \quad . \tag{2.7}$$

The <u>one vector</u> $\underline{e}$ is defined by

$$\underline{e} = \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} = \sum_{i=1}^{n} \underline{e}_i \quad . \tag{2.8}$$

The one matrix $\underline{E}$ is defined by

$$\underline{E} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \cdot & \cdot & \cdot & \cdot \\ 1 & 1 & \dots & 1 \end{bmatrix} = \underline{e}\,\underline{e}' \quad . \tag{2.9}$$

The trace of an $n \times n$ matrix $\underline{A}$ is defined by

$$\mathrm{tr}\,[\underline{A}] = \sum_{i=1}^{n} a_{ii} \quad . \tag{2.10}$$

The trace has the very useful properties

$$\mathrm{tr}[\underline{A} + \underline{B}] = \mathrm{tr}[\underline{A}] + \mathrm{tr}[\underline{B}] \tag{2.11}$$

$$\mathrm{tr}[\underline{A}\,\underline{B}] = \mathrm{tr}[\underline{B}\,\underline{A}] \quad . \tag{2.12}$$

The determinant of an $n \times n$ matrix $\underline{A}$ will be denoted by

$$\det\,[\underline{A}] \ .$$

3. SPACES

We shall denote by

R_n : the set of all real column vectors $\underline{v}$ with n components $v_1, v_2, \dots, v_n$

M_{nn} : the set of all real $n \times n$ matrices .

Both R_n and M_{nn} are linear vector spaces.

The unit vectors $\underline{e}_1, \underline{e}_2, \dots, \underline{e}_n$ (see Eq. (2.2)) belong to R_n and, furthermore, form a basis in R_n. Thus, every $\underline{v} \in R_n$ can be represented by

$$\underline{v} = \sum_{i=1}^{n} v_i \underline{e}_i \quad . \tag{3. 1}$$

Similarly the unit matrices $\underline{E}_{ij}$ belong to M_{nn} and they form a basis in M_{nn}. Thus, every $n \times n$ matrix $\underline{A} \in M_{nn}$ can be represented by (a_{ij} are the elements of $\underline{A}$)

$$\underline{A} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} \underline{E}_{ij} \tag{3. 2}$$

The dimension of R_n is n and the dimension of M_{nn} is n^2.

Note that the transpose $\underline{A}'$ of $\underline{A}$ can be written as

$$\underline{A}' = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ji} \underline{E}_{ij} \tag{3. 3}$$

4. SOME USEFUL DECOMPOSITIONS OF A MATRIX

In this section we shall develop certain formulae relating a matrix, its elements, and its row and column vectors.

If $\underline{A}$ is an $n \times n$ matrix we shall denote its row vectors by $\underline{a}_{1*}, \underline{a}_{2*}, \ldots, \underline{a}_{n*}$ and its column vectors by $\underline{a}_{*1}, \underline{a}_{*2}, \ldots, \underline{a}_{*n}$. Thus

$$\underline{a}_{i*} = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \cdot \\ \cdot \\ \cdot \\ a_{in} \end{bmatrix} \quad ; \quad \underline{a}_{*j} = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \cdot \\ \cdot \\ \cdot \\ a_{nj} \end{bmatrix} \tag{4. 1}$$

and, so,

$$\underline{A} = \left[\begin{array}{c|c|c|c} \uparrow & \uparrow & & \uparrow \\ \underline{a}_{*1} & \underline{a}_{*2} & \cdots & \underline{a}_{*n} \\ \downarrow & \downarrow & & \downarrow \end{array}\right] = \left[\begin{array}{c|c|c|c} & & & \\ \underline{a}_{1*} & \underline{a}_{2*} & \cdots & \underline{a}_{n*} \\ & & & \end{array}\right]' \qquad (4.2)$$

We emphasize that both types of vectors $\underline{a}_{i*}$ and $\underline{a}_{*j}$ are <u>column</u> vectors.

We shall now indicate how one can write the elements a_{ij} and the row and column vectors of a matrix $\underline{A}$ in terms of $\underline{A}$ and in terms of the unit vectors $\underline{e}_i$ (see Eq. (2.2))

$$a_{ij} = \underline{e}'_i \, \underline{A} \, \underline{e}_j = \underline{e}'_j \, \underline{A}' \underline{e}_i \qquad (4.3)$$

$$\underline{a}_{i*} = \underline{A}' \, \underline{e}_i \qquad \text{(the transpose of the i-th row of } \underline{A}\text{)} \qquad (4.4)$$

$$\underline{a}'_{i*} = \underline{e}'_i \, \underline{A} \qquad \text{(the i-th row vector of } \underline{A}\text{)} \qquad (4.5)$$

$$\underline{a}_{*j} = \underline{A} \, \underline{e}_j \qquad \text{(the j-th column vector of } \underline{A}\text{)} . \qquad (4.6)$$

The element a_{ij} can also be generated as follows:

$$a_{ij} = \underline{e}'_i \underline{a}_{*j} \qquad (4.7)$$

or

$$a_{ij} = \underline{e}'_j \underline{a}_{i*} \quad . \qquad (4.8)$$

Next we shall indicate the relation of the row and column vectors of $\underline{A}$ to the elements of $\underline{A}$. From Eqs. (4.3), (4.4), (4.5], and (4.6) we deduce that

$$\underline{a}_{i*} = \sum_{j=1}^{n} a_{ij}\, \underline{e}_j \tag{4.9}$$

$$\underline{a}'_{i*} = \sum_{j=1}^{n} a_{ij}\, \underline{e}'_j \tag{4.10}$$

$$\underline{a}_{*j} = \sum_{i=1}^{n} a_{ij}\, \underline{e}_i \quad . \tag{4.11}$$

The matrix $\underline{A}$ can be generated as follows: from Eqs. (2.5) and (3.2) we have

$$\underline{A} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}\, \underline{e}_i\, \underline{e}'_j = \sum_{i=1}^{n} \sum_{j=1}^{n} \underline{e}_i\, a_{ij}\, \underline{e}'_j \tag{4.12}$$

From Eqs. (4.12), (4.9), (4.10) and (4.11) we obtain

$$\underline{A} = \sum_{i=1}^{n} \underline{e}_i\, \underline{a}'_{i*} \tag{4.13}$$

and

$$\underline{A} = \sum_{j=1}^{n} \underline{a}_{*j}\, \underline{e}'_j \ . \tag{4.14}$$

5. FORMULAE INVOLVING THE UNIT MATRICES $\underline{E}_{ij}$

First of all if we define the Kronecker delta δ_{ij}

$$\delta_{ij} = \begin{cases} +1 & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j \end{cases} \tag{5.1}$$

then we have the relation

$$\underline{e}'_i \underline{e}_j = \underline{e}'_j \underline{e}_i = \delta_{ij} \quad . \tag{5.2}$$

The following two relations relate operations between unit matrices and unit vectors (see Eq. (2.5))

$$\underline{E}_{ij} \underline{e}_k = \underline{e}_i \underline{e}'_j \underline{e}_k = \delta_{jk} \underline{e}_i \tag{5.3}$$

$$\underline{e}'_k \underline{E}_{ij} = \underline{e}'_k \underline{e}_i \underline{e}'_j = \delta_{ki} \underline{e}'_j \quad . \tag{5.4}$$

The following relations relate unit matrices

$$\underline{E}_{ij} \underline{E}_{km} = \underline{e}_i \underline{e}'_j \underline{e}_k \underline{e}'_m = \delta_{jk} \underline{e}_i \underline{e}'_m = \delta_{jk} \underline{E}_{im} . \tag{5.5}$$

It follows that

$$\underline{E}_{ij} \underline{E}_{ij} = \underline{E}^2_{ij} = \delta_{ji} \underline{E}_{ij} = \delta_{ij} \underline{E}_{ij} \tag{5.6}$$

$$\underline{E}_{ij} \underline{E}_{jk} = \delta_{jj} \underline{E}_{ik} = \underline{E}_{ik} \tag{5.7}$$

$$\underline{E}_{ij} \underline{E}_{ji} = \underline{E}_{ii} \tag{5.8}$$

$$\underline{E}^{\alpha}_{ii} = \underline{E}_{ii} \qquad ; \qquad \alpha = 1, 2, \ldots \tag{5.9}$$

$$\underline{E}_{ij} \underline{E}_{jk} \underline{E}_{km} = \underline{E}_{ik} \underline{E}_{km} = \underline{E}_{im} \quad . \tag{5.10}$$

Equation (5.10) generalizes to

$$\underline{E}_{i_1 i_2} \underline{E}_{i_2 i_3} \underline{E}_{i_3 i_4} \ldots \underline{E}_{i_{\beta-1} i_\beta} = \underline{E}_{i_1 i_\beta} \quad . \tag{5.11}$$

We shall next consider the matrix $\underline{E}_{ij}\,\underline{A}$. From Eqs. (3.2), (4.10), and (5.5) we establish that

$$\underline{E}_{ij}\,\underline{A} = \underline{E}_{ij} \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} a_{\alpha\beta}\,\underline{E}_{\alpha\beta} = \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} a_{\alpha\beta}\,\underline{E}_{ij}\,\underline{E}_{\alpha\beta}$$

$$= \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} a_{\alpha\beta}\,\delta_{j\alpha}\,\underline{E}_{i\beta} = \sum_{\beta=1}^{n} a_{j\beta}\,\underline{E}_{i\beta} = \sum_{\beta=1}^{n} a_{j\beta}\,\underline{e}_i\,\underline{e}'_\beta \qquad (5.12)$$

which reduces to (in view of Eq. (4.10))

$$\underline{E}_{ij}\,\underline{A} = \underline{e}_i\,\underline{a}'_{j*} \quad . \qquad (5.13)$$

Similarly we can establish that

$$\underline{A}\,\underline{E}_{ij} = \underline{a}_{*i}\,\underline{e}'_j \qquad (5.14)$$

and that

$$\underline{E}_{ij}\,\underline{A}\,\underline{E}_{km} = a_{jk}\,\underline{E}_{im} \quad . \qquad (5.15)$$

6. INNER PRODUCTS AND THE TRACE FUNCTION

Suppose that $\underline{v}$ and $\underline{w}$ are n-vectors (elements of R_n) ; then the common scalar product

$$(\underline{v}, \underline{w}) = \underline{v}'\underline{w} = \underline{w}'\underline{v} = \sum_{i=1}^{n} v_i w_i \qquad (6.1)$$

is an inner product.

In an analogous manner we define an inner product between two matrices. Let us suppose that $\underline{A}$ and $\underline{B}$, with elements a_{ij} and b_{ij} respectively, are elements of M_{nn}. It can be shown that the mapping

$$(\underline{\ },\underline{B}) = \mathrm{tr}[\underline{A}\,\underline{B}'] = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\, b_{ij} \qquad (6.2)$$

has all the properties of an inner product because

$$\mathrm{tr}[\underline{A}\,\underline{B}'] = \mathrm{tr}[\underline{B}\,\underline{A}'] \qquad (6.3)$$

$$\mathrm{tr}\,[\,\underline{A}\,\underline{B}'] = r\,\mathrm{tr}[\underline{A}\,\underline{B}'] \qquad (\,r : \text{real scalar})r \qquad (6.4)$$

$$\mathrm{tr}[(\underline{A}+\underline{B})\,\underline{C}'] = \mathrm{tr}[\underline{A}\,\underline{C}'] + \mathrm{tr}[\underline{B}\,\underline{C}']\,. \qquad (6.5)$$

We shall present below some interesting properties of the trace. Since

$$\mathrm{tr}[\underline{A}] = \sum_{i=1}^{n} a_{ii} \qquad (6.6)$$

and since (see Eq. (4.3))

$$a_{ii} = \underline{e}'_i\,\underline{A}\,\underline{e}_i, \qquad (6.7)$$

then

$$\mathrm{tr}[\underline{A}\,] = \sum_{i=1}^{n} \underline{e}'_i\,\underline{A}\,\underline{e}_i \quad . \qquad (6.8)$$

From Eqs. (6.8), (4.5), and (4.6) we also obtain

$$\mathrm{tr}[\underline{A}\,] = \sum_{i=1}^{n} \underline{e}'_i\,\underline{a}_{*i} \qquad (6.9)$$

$$\mathrm{tr}[\underline{A}\,] = \sum_{i=1}^{n} \underline{a}'_{i*}\underline{e}_i\,. \qquad (6.10)$$

Now we shall consider $\mathrm{tr}[\underline{A}\,\underline{B}]$. From Eq. (6.8) we have

$$\mathrm{tr}[\underline{A}\,\underline{B}] = \sum_{i=1}^{n} \underline{e}'_i\,\underline{A}\,\underline{B}\,\underline{e}_i \quad . \tag{6.11}$$

We can also express the $\mathrm{tr}[\underline{A}\,\underline{B}]$ in terms of the column and row vectors of $\underline{A}$ and $\underline{B}$. From Eqs. (6.11), (4.5) and (4.6) we have

$$\mathrm{tr}[\underline{A}\,\underline{B}] = \sum_{i=1}^{n} \underline{a}'_{i*}\,\underline{b}_{*i} \,. \tag{6.12}$$

Since (see Eq. (2.12))

$$\mathrm{tr}[\underline{A}\,\underline{B}] = \mathrm{tr}[\underline{B}\,\underline{A}] \tag{6.13}$$

we obtain similarly

$$\mathrm{tr}[\underline{A}\,\underline{B}] = \sum_{i=1}^{n} \underline{b}'_{i*}\,\underline{a}_{*i} \tag{6.14}$$

and that

$$\mathrm{tr}[\underline{A}\,\underline{B}] = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ik}\, b_{ki} \,. \tag{6.15}$$

Similarly we deduce that

$$\mathrm{tr}[\underline{A}\,\underline{B}'] = \sum_{i=1}^{n} \underline{a}'_{i*}\,\underline{b}_{i*} \tag{6.16}$$

$$\mathrm{tr}[\underline{A}\,\underline{B}'] = \sum_{i=1}^{n} \underline{a}'_{*i}\,\underline{b}_{*i} \,. \tag{6.17}$$

Another very interesting formula is the following. Let $\underline{v}$ and $\underline{w}$ be two column vectors; then $\underline{v}\,\underline{w}'$ and $\underline{w}\,\underline{v}'$ are $n \times n$ matrices. Hence, by Eq. (6.8),

$$\mathrm{tr}[\underline{v}\ \underline{w}'] = \sum_{i=1}^{n} \underline{e}_i'\ \underline{v}\ \underline{w}'\ \underline{e}_i \ . \tag{6.18}$$

But

$$\left.\begin{aligned} \underline{e}_i'\ \underline{v} &= v_i \\ \underline{w}'\ \underline{e}_i &= w_i \end{aligned}\right\} \tag{6.19}$$

and, so,

$$\mathrm{tr}[\underline{v}\ \underline{w}'] = \sum_{i=1}^{n} v_i\ w_i = \underline{w}'\ \underline{v} \ . \tag{6.20}$$

Since

$$\mathrm{tr}[\underline{v}\ \underline{w}'] = \underline{w}'\ \underline{v} \tag{6.21}$$

$$\mathrm{tr}[\underline{w}\ \underline{v}'] = \underline{v}'\ \underline{w} \tag{6.22}$$

and,so,

$$\mathrm{tr}[\underline{v}\ \underline{w}'] = \mathrm{tr}[\underline{w}\ \underline{v}'] \ . \tag{6.23}$$

Next we consider $\mathrm{tr}[\underline{A}\ \underline{B}\ \underline{C}]$. From Eq. (6.8) we have

$$\mathrm{tr}[\underline{A}\ \underline{B}\ \underline{C}] = \sum_{i=1}^{n} \underline{e}_i'\ \underline{A}\ \underline{B}\ \underline{C}\ \underline{e}_i \ . \tag{6.24}$$

It follows that

$$\mathrm{tr}[\underline{A}\ \underline{B}\ \underline{C}] = \sum_{i=1}^{n} \underline{a}_{i*}'\ \underline{B}\ \underline{c}_{*i} \ . \tag{6.25}$$

Since

$$\underline{B} = \sum_{j=1}^{n} \underline{e}_j\ \underline{b}_{j*}' \tag{6.26}$$

we can also deduce that

$$\mathrm{tr}[\underline{A}\,\underline{B}\,\underline{C}] = \sum_{i=1}^{n}\sum_{j=1}^{n} \underline{a}'_{i*}\,\underline{c}_j\,\underline{b}'_{j*}\,\underline{c}_{*i} \quad . \tag{6.27}$$

Additional relationships can be derived using the equations

$$\mathrm{tr}[\underline{A}\,\underline{B}\,\underline{C}] = \mathrm{tr}[\underline{B}\,\underline{C}\,\underline{A}] = \mathrm{tr}[\underline{C}\,\underline{A}\,\underline{B}]. \tag{6.28}$$

7. DIFFERENTIALS AND GRADIENT MATRICES

The relations which we have established will be used to develop compact notations for differentiation of matrix quantities.

Let $\underline{x}$ be a column vector with components $x_1, x_2, \ldots, x_n$. Then the differential $d\underline{x}$ of $\underline{x}$ is simply

$$d\underline{x} = \begin{bmatrix} dx_1 \\ dx_2 \\ \cdot \\ \cdot \\ \cdot \\ dx_n \end{bmatrix} . \tag{7.1}$$

Now let $f(\cdot)$ be a scalar real valued function so that

$$f(\underline{x}) \triangleq f(x_1, x_2, \ldots, x_n) .$$

The gradient vector of $f(\cdot)$ with respect to $\underline{x}$ is defined as

$$\frac{\partial f(\underline{x})}{\partial \underline{x}} = \begin{bmatrix} \frac{\partial f(\underline{x})}{\partial x_1} \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial f(\underline{x})}{\partial x_n} \end{bmatrix} . \tag{7.2}$$

For example, suppose that $n = 2$, and that

$$f(\underline{x}) = f(x_1, x_2) = 3x_1^2 + x_1 x_2 + \frac{1}{2} x_2^2 .$$

Then

$$\frac{\partial f(\underline{x})}{\partial \underline{x}} = \begin{bmatrix} 6x_1 + x_2 \\ \\ x_1 + x_2 \end{bmatrix} .$$

Now let $\underline{X}$ be an $n \times n$ matrix with elements x_{ij} $(i, j = 1, 2, \ldots, n)$. The differential $d\underline{X}$ of X is an $n \times n$ matrix such that

$$d\underline{X} = \begin{bmatrix} dx_{11} & dx_{12} & \cdots & dx_{1n} \\ dx_{21} & dx_{22} & \cdots & dx_{2n} \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} . \tag{7.3}$$

Note that the usual rules prevail:

$$d(a\underline{X}) = a\,d\underline{X} \qquad (a: \text{scalar}) \tag{7.4}$$

$$d(\underline{X}+\underline{Y}) = d\underline{X} + d\underline{Y} \tag{7.5}$$

$$d(\underline{X}\,\underline{Y}) = (d\underline{X})\,\underline{Y} + \underline{X}\,(d\underline{Y}) \,. \tag{7.6}$$

From (7.6) we can obtain the useful formula developed below. Suppose that

$$\underline{X} = \underline{Y}^{-1} \tag{7.7}$$

so that

$$\underline{X}\,\underline{Y} = \underline{I} \quad \text{(the identity matrix)} \tag{7.8}$$

and, so,

$$(d\underline{X})\,\underline{Y} + \underline{X}\,(d\underline{Y}) = d\underline{I} = \underline{0} \,. \tag{7.9}$$

It follows that

$$d\underline{X} = -\,\underline{X}\,(d\underline{Y})\,\underline{Y}^{-1} \tag{7.10}$$

and that

$$d(\underline{Y}^{-1}) = -\,\underline{Y}^{-1}(d\underline{Y})\,\underline{Y}^{-1} \quad . \tag{7.11}$$

Next we consider the concept of the gradient matrix. Let $\underline{X}$ be an $n \times n$ matrix with elements x_{ij}. Let $f(\cdot)$ be a scalar, real-valued function of the x_{ij}, i.e.

$$f(\underline{X}) = f(x_{11}, \ldots, x_{1n}, x_{21}, \ldots, x_{2n}, \ldots) . \tag{7.12}$$

We can compute the partial derivatives

$$\frac{\partial f(\underline{X})}{\partial x_{ij}} \quad ; \quad i, j = 1, 2, \ldots, n \,. \tag{7.13}$$

We define an $n \times n$ matrix $\frac{\partial f(\underline{X})}{\partial \underline{X}}$, called the gradient matrix of $f(\underline{X})$ with respect to $\underline{X}$, as the matrix whose ij-th element is given by (7.13). We can use Eq. (4.12) to precisely define the gradient matrix as follows:

$$\frac{\partial f(\underline{X})}{\partial \underline{X}} = \sum_{ij} \underline{e}_i \frac{\partial f(\underline{X})}{\partial x_{ij}} \underline{e}'_j \tag{7.14}$$

or, from Eq. (3.2), to write

$$\frac{\partial f(\underline{X})}{\partial X} = \sum_{ij} \frac{\partial f(\underline{X})}{\partial x_{ij}} \underline{E}_{ij} \quad . \tag{7.15}$$

For example, suppose that $\underline{X}$ is a 2×2 matrix and that

$$f(\underline{X}) = x_{11}^2 x_{21} + x_{21}^3 - x_{11}x_{22}x_{12} + 5x_{21} \quad .$$

Then

$$\frac{\partial f(\underline{X})}{\partial \underline{X}} = \begin{bmatrix} 2x_{11}x_{21} - x_{22}x_{12} & -x_{11}x_{22} \\ x_{11}^2 + 3x_{21}^2 + 5 & -x_{11}x_{12} \end{bmatrix} \quad .$$

Suppose that the elements x_{ij} of $\underline{X}$ represent independent variables, that is

$$\frac{\partial x_{\alpha\beta}}{\partial x_{ij}} = \begin{cases} 1 & \text{if} \quad \alpha = i, \quad \beta = j \\ 0 & \text{otherwise} \end{cases} \quad . \tag{7.16}$$

A useful formula is as follows:

$$\boxed{\frac{d}{dx_{ij}} \underline{X} = \frac{d\underline{X}}{dx_{ij}} = \underline{E}_{ij}} \quad (7.17)$$

If $\underline{X} = \underline{X}'$, i.e. if $\underline{X}$ is symmetric; then $x_{ij} = x_{ji}$ for all i and j. Clearly the differential $d\underline{X}$ is symmetric and

$$d\underline{X} = d\underline{X}' \quad (7.18)$$

$$(d\underline{X})' = d\underline{X} \quad . \quad (7.19)$$

8. GRADIENT MATRICES OF TRACE FUNCTIONS

In this section we shall derive formulae which are useful when one is interested in obtaining the gradient matrix of the trace of a matrix which depends upon the matrix $\underline{X}$. Throughout the section, we shall assume that $\underline{X}$ is an $n \times n$ matrix with elements x_{ij} such that

$$\frac{\partial x_{\alpha\beta}}{\partial x_{ij}} = \begin{cases} 1 & \text{if} \quad \alpha = i \ , \quad \beta = j \\ 0 & \text{otherwise} \ . \end{cases} \quad (8.1)$$

First, we shall compute

$$\frac{\partial}{\partial \underline{X}} \operatorname{tr}[\underline{X}] \quad . \quad (8.2)$$

Since the differential and the trace are linear operators we have

$$d \operatorname{tr}[\underline{X}] = \operatorname{tr}[d\underline{X}] \ . \quad (8.3)$$

Hence, in view of (7.17)

$$\frac{d}{dx_{ij}} \operatorname{tr}[\underline{X}] = \operatorname{tr}\left[\frac{d\underline{X}}{dx_{ij}}\right] = \operatorname{tr}[\underline{E}_{ij}] \ . \quad (8.4)$$

From (7.15) and (8.4) we have

$$\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{X}] = \sum_{ij} \mathrm{tr}[\underline{E}_{ij}] \underline{E}_{ij} . \tag{8.5}$$

But

$$\mathrm{tr}[\underline{E}_{ij}] = \delta_{ij} . \tag{8.6}$$

It follows from (8.5) and (8.6) that

$$\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{X}] = \sum_{ij} \delta_{ij} \underline{E}_{ij} = \sum_{i} \underline{E}_{ii} . \tag{8.7}$$

In view of (2.7) we conclude that

$$\boxed{\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{X}] = \underline{I}} \tag{8.8}$$

Next we shall compute the matrix

$$\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{A}\,\underline{X}] . \tag{8.9}$$

Proceeding as above we have:

$$\frac{\partial}{\partial x_{ij}} \mathrm{tr}[\underline{A}\,\underline{X}] = \mathrm{tr}\left[\underline{A} \frac{d\underline{X}}{dx_{ij}}\right]$$

$$= \mathrm{tr}[\underline{A}\,\underline{E}_{ij}] \qquad \text{(by (7.17))} .$$

But

$$\frac{\partial}{\partial \underline{X}} tr[\underline{A}\,\underline{X}] = \sum_{ij} \underline{e}_{ki} \frac{\partial}{\partial x_{ij}} tr[\underline{A}\,\underline{X}]\underline{e}'_j \qquad \text{(by (7.14))}$$

$$= \sum_{ij} \underline{e}_i\, tr[\underline{A}\,\underline{E}_{ij}]\,\underline{e}'_j \qquad \text{(by (8.10))}$$

$$= \sum_{ijk} \underline{e}_i\, \underline{e}'_k\, \underline{A}\,\underline{E}_{ij}\underline{e}_k\underline{e}'_j \qquad \text{(by (6.8))}$$

$$= \sum_{ijk} \underline{E}_{ik}\, \underline{A}\,\underline{E}_{ij}\underline{E}_{kj} \qquad \text{(by (2.5))}$$

$$= \sum_{ijk} \underline{E}_{ik}\, \underline{A}\, \delta_{jk}\underline{E}_{ij} \qquad \text{(by (5.5))}$$

$$= \sum_{ij} \underline{E}_{ij}\, \underline{A}\, \underline{E}_{ij}$$

$$= \sum_{ij} a_{ji}\, \underline{E}_{ij} \qquad \text{(by (5.15)}$$

$$= \underline{A}' \quad . \qquad \text{(by (3.3))}$$

Thus, we have shown that

$$\boxed{\frac{\partial}{\partial \underline{X}} tr[\underline{A}\,\underline{X}] = \underline{A}'} \quad . \qquad (8.10)$$

In a completely analogous manner we find the following

$$\frac{\partial}{\partial \underline{X}} tr[\underline{A}\,\underline{X}] = \underline{A}' \qquad (8.11)$$

$$\frac{\partial}{\partial \underline{X}} tr[\underline{A}\,\underline{X}'] = \underline{A} \qquad (8.12)$$

$$\frac{\partial}{\partial \underline{X}} \operatorname{tr}[\underline{A}\,\underline{X}\,\underline{B}] = \underline{A}'\underline{B}' \tag{8.13}$$

$$\frac{\partial}{\partial \underline{X}} \operatorname{tr}[\underline{A}\,\underline{X}'\underline{B}] = \underline{B}\,\underline{A} \tag{8.14}$$

$$\frac{\partial}{\partial \underline{X}'} \operatorname{tr}[\underline{A}\,\underline{X}] = \underline{A} \tag{8.15}$$

$$\frac{\partial}{\partial \underline{X}'} \operatorname{tr}[\underline{A}\,\underline{X}'] = \underline{A}' \tag{8.16}$$

$$\frac{\partial}{\partial \underline{X}'} \operatorname{tr}[\underline{A}\,\underline{X}\,\underline{B}] = \underline{B}\,\underline{A} \tag{8.17}$$

$$\frac{\partial}{\partial \underline{X}'} \operatorname{tr}[\underline{A}\,\underline{X}'\underline{B}] = \underline{A}'\underline{B}' \tag{8.18}$$

A useful lemma (which was proved in the derivation of Eq. (8.10)) is the following:

<u>Lemma 8.1</u>

<u>If</u> $\frac{\partial}{\partial x_{ij}} \operatorname{tr}[\underline{A}\,\underline{X}] = \operatorname{tr}[\underline{A}\,\underline{E}_{ij}]$, <u>then</u> $\frac{\partial}{\partial \underline{X}} \operatorname{tr}[\underline{A}\,\underline{X}] = \underline{A}'$.

Next we turn our attention to the derivation of gradient matrices of trace functions involving quadratic forms of the matrix $\underline{X}$.

Consider

$$\frac{\partial}{\partial \underline{X}} \operatorname{tr}[\underline{X}^2] \quad . \tag{8.19}$$

Since

$$d \operatorname{tr}[\underline{X}^2] = \operatorname{tr}[d\underline{X}^2] = \operatorname{tr}[\underline{X}\, d\underline{X} + (d\underline{X})\underline{X}]$$

$$= \operatorname{tr}[\underline{X}\, d\underline{X}] + \operatorname{tr}[(d\underline{X})\underline{X}]$$

$$= \operatorname{tr}[\underline{X}\, d\underline{X}] + \operatorname{tr}[\underline{X}\, d\underline{X}] = 2 \operatorname{tr}[\underline{X}\, d\underline{X}] \tag{8.20}$$

we conclude that

$$\frac{\partial}{\partial x_{ij}} \operatorname{tr}[\underset{\sim}{X}^2] = 2 \operatorname{tr}\left[\underset{\sim}{X} \frac{d\underset{\sim}{X}}{dx_{ij}}\right] = 2 \operatorname{tr}[\underset{\sim}{X}\,\underset{\sim}{E}_{ij}] . \tag{8.21}$$

It follows from Lemma 8.1 that

$$\boxed{\frac{\partial}{\partial \underset{\sim}{X}} \operatorname{tr}[\underset{\sim}{X}^2] = 2\underset{\sim}{X}'} \tag{8.22}$$

In a similar fashion one can prove that

$$\boxed{\frac{\partial}{\partial \underset{\sim}{X}} \operatorname{tr}[\underset{\sim}{X}\,\underset{\sim}{X}'] = 2\underset{\sim}{X}} \tag{8.23}$$

Next we consider

$$\frac{\partial}{\partial \underset{\sim}{X}} \operatorname{tr}[\underset{\sim}{A}\,\underset{\sim}{X}\,\underset{\sim}{B}\,\underset{\sim}{X}] . \tag{8.24}$$

Since

$$\begin{aligned} d \operatorname{tr}[\underset{\sim}{A}\,\underset{\sim}{X}\,\underset{\sim}{B}\,\underset{\sim}{X}] &= \operatorname{tr}[d(\underset{\sim}{A}\,\underset{\sim}{X}\,\underset{\sim}{B}\,\underset{\sim}{X})] \\ &= \operatorname{tr}[\underset{\sim}{A}(d\underset{\sim}{X})\underset{\sim}{B}\,\underset{\sim}{X}] + \operatorname{tr}[\underset{\sim}{A}\,\underset{\sim}{X}\,\underset{\sim}{B}(d\underset{\sim}{X})] \\ &= \operatorname{tr}[\underset{\sim}{B}\,\underset{\sim}{X}\,\underset{\sim}{A}(d\underset{\sim}{X})] + \operatorname{tr}[\underset{\sim}{A}\,\underset{\sim}{X}\,\underset{\sim}{B}(d\underset{\sim}{X})] \\ &= \operatorname{tr}[(\underset{\sim}{B}\,\underset{\sim}{X}\,\underset{\sim}{A} + \underset{\sim}{A}\,\underset{\sim}{X}\,\underset{\sim}{B})(d\underset{\sim}{X})] \end{aligned} \tag{8.25}$$

we conclude that

$$\boxed{\frac{\partial}{\partial \underset{\sim}{X}} \operatorname{tr}[\underset{\sim}{A}\,\underset{\sim}{X}\,\underset{\sim}{B}\,\underset{\sim}{X}] = \underset{\sim}{A}'\underset{\sim}{X}'\underset{\sim}{B}' + \underset{\sim}{B}'\underset{\sim}{X}'\underset{\sim}{A}'} \tag{8.26}$$

Next we consider

$$\frac{\partial}{\partial \underline{X}} \operatorname{tr}[\underline{A}\,\underline{X}\,\underline{B}\,\underline{X}'] \,. \tag{8.27}$$

Since

$$d\operatorname{tr}[\underline{A}\,\underline{X}\,\underline{B}\,\underline{X}'] = \operatorname{tr}[\underline{A}(d\underline{X})\underline{B}\,\underline{X}'] + \operatorname{tr}[\underline{A}\,\underline{X}\,\underline{B}(d\underline{X}')]$$

$$= \operatorname{tr}[\underline{B}\,\underline{X}'\underline{A}(d\underline{X})] + \operatorname{tr}[(d\underline{X})'\,\underline{A}\,\underline{X}\,\underline{B}]$$

$$= \operatorname{tr}[\underline{B}\,\underline{X}'\underline{A}(d\underline{X})] + \operatorname{tr}[\underline{B}'\underline{X}'\underline{A}'(d\underline{X})]$$

$$= \operatorname{tr}[(\underline{B}\,\underline{X}'\underline{A} + \underline{B}'\underline{X}'\underline{A}')(d\underline{X})] \tag{8.28}$$

(because $(d\underline{X}') = (d\underline{X})'$ and because $\operatorname{tr}[\underline{Y}] = \operatorname{tr}[\underline{Y}']$ for all $\underline{Y}$), it follows that

$$\boxed{\frac{\partial}{\partial \underline{X}} \operatorname{tr}[\underline{A}\,\underline{X}\,\underline{B}\,\underline{X}'] = \underline{A}'\underline{X}\,\underline{B}' + \underline{A}\,\underline{X}\,\underline{B}} \tag{8.29}$$

The following two equations involve higher powers of $\underline{X}$ and they are easy to derive

$$\frac{\partial}{\partial \underline{X}} \operatorname{tr}[\underline{X}^n] = n(\underline{X}')^{n-1} = n(\underline{X}^{n-1})' \tag{8.30}$$

$$\frac{\partial}{\partial \underline{X}} \operatorname{tr}[\underline{A}\,\underline{X}^n] = (\underline{A}\,\underline{X}^{n-1} + \underline{X}\,\underline{A}\,\underline{X}^{n-2} + \underline{X}^2\,\underline{A}\,\underline{X}^{n-3} + \cdots + \underline{X}^{n-2}\underline{A}\,\underline{X} + \underline{X}^{n-1}\,\underline{A})' \tag{8.31}$$

Equation (8.31) can also be written as

$$\frac{\partial}{\partial \underline{X}} \operatorname{tr}[\underline{A}\,\underline{X}^n] = \left(\sum_{i=0}^{n-1} \underline{X}^i\,\underline{A}\,\underline{X}^{n-1-i}\right)' \tag{8.32}$$

The two formulae above provide us with the capability of solving for the gradient matrices of trace functions of polynomials in $\underline{X}$. A particular function of interest is the exponential matrix function $e^{\underline{X}}$ which is commonly defined by the infinite series

$$e^{\underline{X}} = \underline{I} + \underline{X} + \frac{1}{2!}\underline{X}^2 + \frac{1}{3!}\underline{X}^3 + \cdots = \sum_{i=0}^{\infty} \frac{1}{i!}\underline{X}^i \quad . \tag{8.33}$$

We proceed to evaluate

$$\frac{\partial}{\partial \underline{X}} \mathrm{tr}[e^{\underline{X}}] \quad . \tag{8.34}$$

Since

$$\mathrm{tr}[e^{\underline{X}}] = \mathrm{tr}\left[\sum_{i=0}^{\infty} \frac{1}{i!}\underline{X}^i\right] = \sum_{i=0}^{\infty} \frac{1}{i!}\mathrm{tr}[\underline{X}^i] \tag{8.35}$$

we can use Eq.(8.30) to find that

$$\boxed{\frac{\partial}{\partial \underline{X}} \mathrm{tr}[e^{\underline{X}}] = e^{\underline{X}}} \quad . \tag{8.36}$$

We shall next compute

$$\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{X}^{-1}] \quad . \tag{8.37}$$

First recall the relation (see Eq. (7.11))

$$d\underline{X}^{-1} = -\underline{X}^{-1}(d\underline{X})\underline{X}^{-1} \quad . \tag{8.38}$$

It follows that

$$d\,\mathrm{tr}[\underline{X}^{-1}] = \mathrm{tr}[d\underline{X}^{-1}] = -\mathrm{tr}[\underline{X}^{-1}(d\underline{X})\underline{X}^{-1}] \tag{8.39}$$

and, so,

$$\frac{\partial}{\partial x_{ij}} \mathrm{tr}[\underline{X}^{-1}] = -\mathrm{tr}\left[\underline{X}^{-1}\frac{d\underline{X}}{dx_{ij}}\underline{X}^{-1}\right]$$

$$= -\mathrm{tr}[\underline{X}^{-1}\underline{E}_{ij}\underline{X}^{-1}]$$

$$= -\mathrm{tr}[\underline{X}^{-2}\underline{E}_{ij}]\,. \tag{8.40}$$

From Eq. (8.40) and Lemma 8.1 we conclude that

$$\boxed{\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{X}^{-1}] = -(\underline{X}^{-2})'} \tag{8.41}$$

In a similar fashion we can show that

$$\boxed{\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{A}\,\underline{X}^{-1}\underline{B}] = -(\underline{X}^{-1}\underline{B}\,\underline{A}\,\underline{X}^{-1})'} \tag{8.42}$$

9. GRADIENT MATRICES OF DETERMINANT FUNCTIONS

The trace $\mathrm{tr}[\underline{X}]$ and the determinant $\det[\underline{X}]$ of a matrix $\underline{X}$ are the two most used scalar functions of a matrix. In the previous section we developed relations for the gradient matrix of trace functions. In this section we shall develop similar relations for the gradient matrix of determinant functions.

Before commencing the computations it is necessary to state some of the properties of the determinant function. Let $\underline{X}$ be an $n \times n$ matrix. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of $\underline{X}$; for simplicity we shall assume that these eigenvalues are distinct. It is always true that the trace of $\underline{X}$ equals to the sum of the eigenvalues while the determinant of $\underline{X}$ is the product of the eigenvalues; in other words,

$$\mathrm{tr}[\underline{X}] = \lambda_1 + \lambda_2 + \cdots + \lambda_n \tag{9.1}$$

$$\det[\underline{X}] = \lambda_1 \lambda_2 \cdots \lambda_n \quad . \tag{9.2}$$

The determinant has the following properties:

$$\det[\underline{X}\,\underline{Y}] = \det[\underline{X}]\,\det[\underline{Y}] \tag{9.3}$$

$$\det[\underline{X} + \underline{Y}] \neq \det[\underline{X}] + \det[\underline{Y}] \tag{9.4}$$

$$\det[\underline{I}] = 1 \tag{9.5}$$

$$\det[\underline{X}^{-1}] = 1/\det[\underline{X}] \tag{9.6}$$

$$\det[\underline{X}^n] = (\det \underline{X})^n \tag{9.7}$$

$$\det[\underline{X}] = \det[\underline{X}'] \quad . \tag{9.8}$$

In this section we shall use $\underline{\Lambda}$ to denote the diagonal matrix, whose diagonal is formed by the eigenvalues of $\underline{X}$, i.e.

$$\underline{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \quad . \tag{9.9}$$

Clearly

$$\mathrm{tr}[\underline{\Lambda}] = \lambda_1 + \lambda_2 + \cdots + \lambda_n \tag{9.10}$$

$$\det[\underline{\Lambda}] = \lambda_1 \lambda_2 \cdots \lambda_n \tag{9.11}$$

and, so,

$$\mathrm{tr}[\underline{X}] = \mathrm{tr}[\underline{\Lambda}] \tag{9. 12}$$

$$\det[\underline{X}] = \det[\underline{\Lambda}]. \tag{9. 13}$$

Using the differential operator we have

$$d(\mathrm{tr}[\underline{X}]) = d(\mathrm{tr}[\underline{\Lambda}]) \tag{9. 14}$$

$$d(\det[\underline{X}]) = d(\det[\underline{\Lambda}]) . \tag{9. 15}$$

Now we compute the differential of $\det[\underline{X}]$ (provided $\underline{X}$ is nonsingular)

$$\begin{aligned} d(\det[\underline{X}]) &= d(\lambda_1 \lambda_2 \ldots \lambda_n) \\ &= (d\lambda_1)\lambda_2\lambda_3\ldots\lambda_n + \lambda_1(d\lambda_2)\lambda_3\ldots\lambda_n \\ &+ \cdots + \lambda_1\lambda_2\ldots\lambda_{n-1}(d\lambda_n) \\ &= (\det[\underline{\Lambda}])\left[\frac{d\lambda_1}{\lambda_1} + \frac{d\lambda_2}{\lambda_2} + \cdots + \frac{d\lambda_n}{\lambda_n}\right] . \end{aligned} \tag{9. 16}$$

We note that we can identify

$$\sum_{i=1}^{n} \frac{d\lambda_i}{\lambda_i} = \mathrm{tr}[\underline{\Lambda}^{-1} d\,\underline{\Lambda}] \quad \text{and, so,}$$

in view of (9. 13) we have

$$d(\det[\underline{X}]) = (\det[\underline{X}])\,\mathrm{tr}[\underline{\Lambda}^{-1} d\,\underline{\Lambda}] . \tag{9. 17}$$

We shall now prove the following lemma:

Lemma 9.1 If $\underline{X}$ is nonsingular and if it has distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ then

$$\mathrm{tr}[\underline{\Lambda}^{-1} d\underline{\Lambda}] = \mathrm{tr}[\underline{X}^{-1} d\underline{X}] \tag{9.18}$$

Proof: $\underline{X}$ and $\underline{\Lambda}$ are related by the similarity transformation

$$\underline{\Lambda} = \underline{P}^{-1} \underline{X}\, \underline{P} \,. \tag{9.19}$$

Thus

$$\underline{\Lambda}^{-1} = \underline{P}^{-1} \underline{X}^{-1} \underline{P} \,. \tag{9.20}$$

From Eq. (9.19) we have

$$\underline{P}\, \underline{\Lambda} = \underline{X}\, \underline{P} \tag{9.21}$$

and, so,

$$(d\underline{P})\underline{\Lambda} + \underline{P}(d\underline{\Lambda}) = (d\underline{X})\underline{P} + \underline{X}(d\underline{P}) \,. \tag{9.22}$$

It follows that

$$d\underline{\Lambda} = \underline{P}^{-1}(d\underline{X})\underline{P} + \underline{P}^{-1}\underline{X}(d\underline{P}) - \underline{P}^{-1}(d\underline{P})\underline{\Lambda} \,. \tag{9.23}$$

From Eqs. (9.12), (9.20) and (9.23) we obtain

$$\begin{aligned} \underline{\Lambda}^{-1} d\underline{\Lambda} &= \underline{P}^{-1}\underline{X}^{-1}\underline{P}\,\underline{P}^{-1}(d\underline{X})\underline{P} + \underline{P}^{-1}\underline{X}^{-1}\underline{P}\,\underline{P}^{-1}\underline{X}(d\underline{P}) \\ &\quad - \underline{P}^{-1}\underline{X}^{-1}\underline{P}\,\underline{P}^{-1}(d\underline{P})\underline{P}^{-1}\underline{X}\,\underline{P} \\ &= \underline{P}^{-1}\underline{X}^{-1}(d\underline{X})\underline{P} + \underline{P}^{-1}(d\underline{P}) - \underline{P}^{-1}\underline{X}^{-1}(d\underline{P})\underline{P}^{-1}\underline{X}\,\underline{P} \,. \end{aligned} \tag{9.24}$$

Forming the trace of both sides and using the properties (2. 11) and (2. 12) we find

$$\mathrm{tr}[\underline{\Lambda}^{-1} d\underline{\Lambda}] = \mathrm{tr}[\underline{X}^{-1} d\underline{X}] \tag{9.25}$$

Q. E. D.

Using Eqs. (9. 18) and (9. 19) we arrive at[†]

$$d(\det[\underline{X}]) = (\det[\underline{X}])\,\mathrm{tr}[\underline{X}^{-1} d\underline{X}]\,. \tag{9.26}$$

We can now compute the gradient matrix of $\det[\underline{X}]$, i.e. the matrix

$$\frac{\partial}{\partial \underline{X}} \det[\underline{X}]\,. \tag{9.27}$$

From (9. 26) we have

$$\begin{aligned} \frac{\partial}{\partial x_{ij}} \det[\underline{X}] &= (\det[\underline{X}])\,\mathrm{tr}\left[\underline{X}^{-1} \frac{d\underline{X}}{dx_{ij}}\right] \\ &= \det[\underline{X}]\ \mathrm{tr}[\underline{X}^{-1} \underline{E}_{ij}] \end{aligned} \tag{9.28}$$

and, so, Lemma 8. 1 yields

$$\boxed{\frac{\partial}{\partial \underline{X}} \det[\underline{X}] = (\det[\underline{X}])\,(\underline{X}^{-1})'} \tag{9.29}$$

If we write Eq. (9. 26) in the more suggestive form

$$\frac{d(\det[\underline{X}])}{\det[\underline{X}]} = \mathrm{tr}[\underline{X}^{-1}\, d\underline{X}] \tag{9.30}$$

† Equation (9. 26) is true even if the eigenvalues are not distinct; see Ref. [7], p. 35.

we can see that

$$d(\log \det[\underline{X}]) = \mathrm{tr}[\underline{X}^{-1} d\underline{X}] \qquad (9.31)$$

and that

$$\boxed{\frac{\partial}{\partial \underline{X}} \log \det[\underline{X}] = (\underline{X}^{-1})'} \qquad (9.32)$$

(a most useful relation). Using the property (9.3) of the determinant function it is easy to prove that

$$\boxed{\frac{\partial}{\partial \underline{X}} \det[\underline{A}\,\underline{X}\,\underline{B}] = (\det[\underline{A}\,\underline{X}\,\underline{B}])(\underline{X}^{-1})'} \qquad (9.33)$$

Also, it is easy to show (in view of (9.8)) that

$$\boxed{\frac{\partial}{\partial \underline{X}} \det[\underline{X}'] = \frac{\partial}{\partial \underline{X}} \det[\underline{X}]} \qquad (9.34)$$

From the obvious relation

$$d(\det[\underline{X}^n] = d(\det[\underline{X}])^n = n(\det[\underline{X}])^{n-1} d(\det \underline{X}) \qquad (9.35)$$

we conclude that Eq. (9.26) yields

$$d(\det[\underline{X}^n]) = n(\det[\underline{X}])^n \mathrm{tr}[\underline{X}^{-1} d\underline{X}] \qquad (9.36)$$

and so

$$\boxed{\frac{\partial}{\partial \underline{X}} \det[\underline{X}^n] = n(\det[\underline{X}])^n (\underline{X}^{-1})'} \qquad (9.37)$$

10. PARTITIONED MATRICES

It is often necessary to work with partitioned matrices. The following formulae are very useful.

Consider the $n \times n$ matrix $\underline{X}$ partitioned as follows:

$$\underline{X} = \left[\begin{array}{c|c} \underline{X}_{11} & \underline{X}_{12} \\ \hline \underline{X}_{21} & \underline{X}_{22} \end{array}\right] \tag{10.1}$$

where

$\underline{X}_{11}$ is $n_1 \times n_1$ matrix

$\underline{X}_{12}$ is $n_1 \times n_2$ matrix

$\underline{X}_{21}$ is $n_2 \times n_1$ matrix

$\underline{X}_{22}$ is $n_2 \times n_2$ matrix

$n_1 + n_2 = n$.

Assume the necessary inverses exist; and that $\underline{X}^{-1}$ is also partitioned as in (10.1). Then

$$\underline{X}^{-1} = \left[\begin{array}{c|c} \underline{X}_{11}^{-1} + \underline{X}_{11}^{-1}\underline{X}_{12}\underline{\Delta}^{-1}\underline{X}_{21}\underline{X}_{11}^{-1} & -\underline{X}_{11}^{-1}\underline{X}_{12}\underline{\Delta}^{-1} \\ \hline -\underline{\Delta}^{-1}\underline{X}_{21}\underline{X}_{11}^{-1} & \underline{\Delta}^{-1} \end{array}\right] \tag{10.2}$$

where

$$\underline{\Delta} = \underline{X}_{22} - \underline{X}_{21}\underline{X}_{11}^{-1}\underline{X}_{12} \quad . \tag{10.3}$$

From (10.2), the following is obtained.

$$(\underline{X}_{11} - \underline{X}_{12}\underline{X}_{22}^{-1}\underline{X}_{21})^{-1} = \underline{X}_{11}^{-1} + \underline{X}_{11}^{-1}\underline{X}_{12}(\underline{X}_{22} - \underline{X}_{21}\underline{X}_{11}^{-1}\underline{X}_{12})^{-1}\underline{X}_{21}\underline{X}_{11}^{-1} \qquad (10.4)$$

with the special case

$$(\underline{I} + \underline{X})^{-1} = \underline{I} - (\underline{I} + \underline{X}^{-1})^{-1} . \qquad (10.5)$$

Other useful formulae are

$$\det[\underline{X}] = \det[\underline{X}_{11} - \underline{X}_{12}\underline{X}_{22}^{-1}\underline{X}_{21}]\det[\underline{X}_{22}] \qquad (10.6)$$

$$\mathrm{tr}[\underline{X}] = \mathrm{tr}[\underline{X}_{11}] + \mathrm{tr}[\underline{X}_{22}] . \qquad (10.7)$$

If $\underline{Y}$ is also partitioned as in (10.1), then

$$\underline{X}\,\underline{Y} = \left[\begin{array}{c|c} \underline{X}_{11}\underline{Y}_{11} + \underline{X}_{12}\underline{Y}_{21} & \underline{X}_{11}\underline{Y}_{12} + \underline{X}_{12}\underline{Y}_{22} \\ \hline \underline{X}_{21}\underline{Y}_{11} + \underline{X}_{22}\underline{Y}_{21} & \underline{X}_{21}\underline{Y}_{12} + \underline{X}_{22}\underline{Y}_{22} \end{array}\right] . \qquad (10.8)$$

TABLE OF GRADIENTS

1. $\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{X}] = \underline{I}$

2. $\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{A}\,\underline{X}] = \underline{A}'$

3. $\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{A}\,\underline{X}'] = \underline{A}$

4. $\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{A}\,\underline{X}\,\underline{B}] = \underline{A}'\underline{B}'$

5. $\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{A}\,\underline{X}'\underline{B}] = \underline{B}\,\underline{A}$

6. $\frac{\partial}{\partial \underline{X}'} \mathrm{tr}[\underline{A}\,\underline{X}] = \underline{A}$

7. $\frac{\partial}{\partial \underline{X}'} \mathrm{tr}[\underline{A}\,\underline{X}'] = \underline{A}'$

8. $\frac{\partial}{\partial \underline{X}'} \mathrm{tr}[\underline{A}\,\underline{X}\,\underline{B}] = \underline{B}\,\underline{A}$

9. $\frac{\partial}{\partial \underline{X}'} \mathrm{tr}[\underline{A}\,\underline{X}'\underline{B}] = \underline{A}'\underline{B}'$

10. $\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{X}\,\underline{X}] = 2\underline{X}'$

11. $\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{X}\,\underline{X}'] = 2\underline{X}$

12. $\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{X}^n] = n(\underline{X}^{n-1})'$

TABLE OF GRADIENTS (Continued)

13. $\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{A}\,\underline{X}^n] = \left(\sum_{i=0}^{n-1} \underline{X}^i \underline{A}\,\underline{X}^{n-1-i} \right)'$

14. $\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{A}\,\underline{X}\,\underline{B}\,\underline{X}] = \underline{A}'\underline{X}'\underline{B}' + \underline{B}'\underline{X}'\underline{A}'$

15. $\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{A}\,\underline{X}\,\underline{B}\,\underline{X}'] = \underline{A}'\underline{X}\,\underline{B}' + \underline{A}\,\underline{X}\,\underline{B}$

16. $\frac{\partial}{\partial \underline{X}} \mathrm{tr}[e^{\underline{X}}] = e^{\underline{X}}$

17. $\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{X}^{-1}] = -(\underline{X}^{-1}\underline{X}^{-1})' = -(\underline{X}^{-2})'$

18. $\frac{\partial}{\partial \underline{X}} \mathrm{tr}[\underline{A}\,\underline{X}^{-1}\underline{B}] = -(\underline{X}^{-1}\underline{B}\,\underline{A}\,\underline{X}^{-1})'$

19. $\frac{\partial}{\partial \underline{X}} \det[\underline{X}] = (\det[\underline{X}])(\underline{X}^{-1})'$

20. $\frac{\partial}{\partial \underline{X}} \log\det[\underline{X}] = (\underline{X}^{-1})'$

21. $\frac{\partial}{\partial \underline{X}} \det[\underline{A}\,\underline{X}\,\underline{B}] = (\det[\underline{A}\,\underline{X}\,\underline{B}])(\underline{X}^{-1})'$

22. $\frac{\partial}{\partial \underline{X}} \det[\underline{X}'] = \frac{\partial}{\partial X} \det[\underline{X}] = (\det[\underline{X}])(\underline{X}^{-1})'$

23. $\frac{\partial}{\partial \underline{X}} \det[\underline{X}^n] = n(\det[\underline{X}])^n (\underline{X}^{-1})'$

REFERENCES

[1] F. C. Schweppe, "Optimization of Signals," Group Report 1964-4, MIT Lincoln Laboratory (1964).

[2] F. C. Schweppe, "On the Distance Between Gaussian Processes: The State Space Approach," JA-2639, MIT Lincoln Laboratory, 1965.

[3] F. C. Schweppe, "Radar Frequency Modulations for Accelerating Targets Under a Bandwidth Constraint," IEEE Trans. on Military Electronics, MIL-9, pp 25-32 (1965).

[4] M. Athans and F. C. Schweppe, "On the Design of Optimal Modulation Schemes via Control Theoretic Concepts, I: Formulation," JA-2634, MIT Lincoln Laboratory, 1965.

[5] M. Athans, "Solution of the Matrix Equation $\frac{d}{dt}$ (t) = A(t) (t) + (t) B(t) + U(t)," Technical Note 1965-26, MIT Lincoln Laboratory, Lexington, Mass., (1965).

[6] F. C. Schweppe and D. Gray, "Radar Signal Design Subject to Simultaneous Peak and Average Power Constraints," JA-2559, MIT Lincoln Laboratory, to be published in IEEE Trans. on Information Theory.

[7] E. Bodewig, Matrix Calculus, North Holland Publishing Company and Interscience Publishers, N. Y., (1956).

[8] R. Bellman, Introduction to Matrix Analysis, McGraw-Hill Book Co., N. Y. (1960).