JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematical Sciences
Vol. 75B, Nos. 3 and 4, July–December 1971

Some Elementary Formulas in "Matrix Calculus" and Their Applications

Jeffrey T. Fong

Institute for Basic Standards, National Bureau of Standards, Washington, D.C. 20234

(March 8, 1971)

A collection of elementary formulas for calculating the gradients of scalar- and matrix-valued functions of one matrix argument is presented. Using some of the well-known properties of the operator "trace" on square matrices, alternative definitions of gradients and simple examples of calculating them using the product rule and the chain rule for differentiation are treated in an expository fashion in both component and matrix notations with emphasis on the latter. Two examples in continuum mechanics are presented to illustrate the application of the so-called "matrix calculus" of differentiable functions.

Key words: Chain rule; continuum mechanics; gradient; matrices; matrix calculus; partial differentiation; product rule; tensor function; trace.

1. Introduction

This is an expository article on the use of matrix notation in the elementary calculus of differentiable functions whose arguments are square matrices. For example, in continuum physics, it is often necessary to work with partial derivatives of a class of functions whose arguments are elements of a square matrix and whose values can be either scalars or square matrices of the same order. Following the notation and basic concepts of tensor functions as treated by Truesdell and Noll [1, pp. 20–35][1], we present here an elementary introduction to the proper formulation of the chain rule and the product rule for differentiation in matrix notation and we include examples, formulas and applications to illustrate the two rules.

The reader is assumed to be familiar with the notions of the trace and the determinant of a matrix $\underline{\underline{A}}=(A_{ij})$, i, $j=1, 2, \ldots, n$, i.e., $\operatorname{tr} \underline{\underline{A}}=\sum_{i=1}^{n} A_{ii}$ and $\det \underline{\underline{A}}=\Sigma\,(-1)^h A_{1\sigma_1}A_{2\sigma_2}\ldots A_{n\sigma_n}$, where the last summation is made over all permutations of $\sigma_1, \sigma_2, \ldots, \sigma_n$, and h is the number of interchanges required to restore the natural order.[2] In particular, the following properties of the operator "trace" are applied frequently throughout the paper:

(a) $\operatorname{tr}(\underline{\underline{A}}+\underline{\underline{B}})=\operatorname{tr}\underline{\underline{A}}+\operatorname{tr}\underline{\underline{B}}$;
(b) $\operatorname{tr}(\underline{\underline{A}}\,\underline{\underline{B}})=\operatorname{tr}(\underline{\underline{B}}\,\underline{\underline{A}})$;
(c) $\operatorname{tr}(\underline{\underline{A}}^T)=\operatorname{tr}\underline{\underline{A}}$, where $\underline{\underline{A}}^T$ denotes the transpose of $\underline{\underline{A}}$;
(d) $\underline{\underline{A}}=\underline{\underline{B}}$, if, and only if, $\operatorname{tr}(\underline{\underline{A}}\,\underline{\underline{C}})=\operatorname{tr}(\underline{\underline{B}}\,\underline{\underline{C}})$ for arbitrary matrix $\underline{\underline{C}}$.

AMS Subject Classification: Primary 15, 88; Secondary 69.

[1] Figures in brackets indicate the literature references at the end of this paper.

[2] A square matrix is denoted by a symbol underlined with two bars indicating the need for two indices in component notation. In general, any quantity with, say, k indices in component notation will be underlined with k bars when the indices are suppressed. For ease of printing, this convention is followed in equations but ignored in text.

2. Gradient of a Scalar Function of a Matrix Argument

Let $\epsilon \equiv \hat{\epsilon}(A_{11}, A_{12}, \ldots, A_{1n}, A_{21}, A_{22}, \ldots, A_{2n}, \ldots, A_{n1}, A_{n2}, \ldots, A_{nn})$ define a scalar-valued function $\hat{\epsilon}$ of n^2 variables A_{km}, $k, m = 1, 2, \ldots, n$, such that the set of variables A_{km} corresponds to the set of components of a square matrix A of order n. In matrix notation, the definition of the scalar function $\hat{\epsilon}$ assumes the following simple form:

$$\epsilon \equiv \hat{\epsilon}(\underline{\underline{A}}). \tag{2.1}$$

If $\hat{\epsilon}$ is differentiable with respect to each variable A_{km}, the set of first partial derivatives of $\hat{\epsilon}$, i.e., $\{D_{km}\hat{\epsilon}, k, m = 1, 2, \ldots, n\}$, can be defined as a matrix-valued function to be denoted by $\nabla\hat{\epsilon}$ where the element $(\nabla\hat{\epsilon})_{km}$ at the kth row and the mth column of $\nabla\hat{\epsilon}$ is given precisely by $D_{km}\hat{\epsilon}$. Let ϵ_A denote the value of the function $\nabla\hat{\epsilon}$ for a given A, then the definition of the function $\nabla\hat{\epsilon}$, to be known as the gradient of $\hat{\epsilon}$, can be stated in both component and matrix notations as follows:

$$\underline{\underline{\epsilon_A}} \equiv [(\epsilon_A)_{km}] \equiv [D_{km}\hat{\epsilon}(A_{pq})] \equiv [(\nabla\hat{\epsilon})_{km}(A_{pq})] \equiv \underline{\underline{\nabla}}\hat{\epsilon}(\underline{\underline{A}}). \tag{2.2}$$

For brevity, we omit the statement that all indices $k, m, p, q, \ldots$, etc., range from 1 to n.

For the purpose of applying those properties of the operator "trace" as listed in the last section, Truesdell and Noll [1] presented an alternative definition of the gradient of a scalar function of a matrix argument as follows:

$$\operatorname{tr}\{\underline{\underline{\epsilon_A^T}}\,\underline{\underline{C}}\} \equiv \frac{d}{ds}\hat{\epsilon}(\underline{\underline{A}} + s\underline{\underline{C}})\,|_{s=0}, \tag{2.3a}$$

or, in component notation,

$$\sum_{k=1}^{n}\sum_{m=1}^{n}(\epsilon_A)_{km}C_{km} \equiv \frac{d}{ds}\hat{\epsilon}(A_{pq} + sC_{pq})\,|_{s=0}, \tag{2.3b}$$

where C, with components C_{pq}, is an arbitrary matrix of the same order as the matrix A. To see that (2.2) and (2.3) are equivalent, we apply the chain rule for differentiation to the expression $\hat{\epsilon}(A_{pq} + sC_{pq})$:

$$\frac{d}{ds}\hat{\epsilon}(A_{pq} + sC_{pq}) = \sum_{k=1}^{n}\sum_{m=1}^{n}[D_{km}\hat{\epsilon}(A_{pq} + sC_{pq})]\left[\frac{d}{ds}(A_{km} + sC_{km})\right]$$

$$= \sum_{k=1}^{n}\sum_{m=1}^{n}\left[D_{km}\hat{\epsilon}\,(A_{pq} + s\,C_{pq})\right]C_{km}. \tag{2.4}$$

If we substitute zero for s in (2.4) and apply (2.2), we obtain (2.3b). Conversely, (2.3b) and the chain rule imply (2.2). The reader may wish to verify that (2.3) indeed defines a unique matrix ϵ_A as a result of the linearity of the operator "trace" and the arbitrariness of the matrix C.

Example 1: $\epsilon = \hat{\epsilon}(\underline{\underline{A}}) = \det \underline{\underline{A}}$.

To calculate the gradient of $\hat{\epsilon}$, we apply the Laplace development of a determinant, i.e.,

$$\det \underline{\underline{A}} = \sum_{k=1}^{n} A_{km}A^{km},\ m \text{ being fixed and not summed,} \tag{2.5}$$

where A^{km} denotes the cofactor of A_{km} defined as $(-1)^{k+m}$ times the complementary minor of A_{km}. Using (2.2) and the fact that the cofactor A^{km} is independent of the element A_{km} of the matrix A, we obtain:

$$\underline{\underline{\epsilon_A}} = [(\epsilon_A)_{km}] = [D_{km}(\det \underline{\underline{A}})] = [A^{km}] = \underline{\underline{A}}^{\text{cof}} \tag{2.6}$$

where A^{cof} denotes the cofactor matrix of A which, by definition, equals the transpose of the adjoint matrix of A. Let us verify the result given in (2.6) by applying the alternative definition of ϵ_A as given in (2.3):

$$\begin{aligned}\operatorname{tr}\{\underline{\underline{\epsilon}}_A^T \underline{\underline{C}}\} &= \frac{d}{ds}\det(\underline{\underline{A}} + s\underline{\underline{C}})\,|_{s=0} \\ &= \frac{d}{ds}\det[\underline{\underline{A}}(\underline{\underline{1}} + s\underline{\underline{A}}^{-1}\underline{\underline{C}})]\,|_{s=0} \\ &= (\det \underline{\underline{A}})\frac{d}{ds}\det(\underline{\underline{1}} + s\underline{\underline{A}}^{-1}\underline{\underline{C}})\,|_{s=0}.\end{aligned} \tag{2.7}$$

Following Truesdell and Noll [1], we introduce another expansion of a determinant:

$$\det(1 + s\underline{\underline{B}}) = 1 + I_1(\underline{\underline{B}})s + I_2(\underline{\underline{B}})s^2 + \ . \ . \ . + I_n(\underline{\underline{B}})s^n, \tag{2.8}$$

where B is any square matrix of order n and $I_1(B), I_2(B), \ldots, I_n(B)$ are the so-called principal invariants of B.[3] In our case, we are only interested in the first principal invariant $I_1(B)$ which equals tr (B). Combining (2.7) and (2.8), we obtain:

$$\begin{aligned}\operatorname{tr}\{\underline{\underline{\epsilon}}_A^T \underline{\underline{C}}\} &= (\det \underline{\underline{A}})\operatorname{tr}(\underline{\underline{A}}^{-1}\underline{\underline{C}}) \\ &= \operatorname{tr}\{(\det \underline{\underline{A}})\underline{\underline{A}}^{-1}\underline{\underline{C}}\},\end{aligned}$$

i.e.,

$$\underline{\underline{\epsilon_A}} = (\det \underline{\underline{A}})(\underline{\underline{A}}^{-1})^T. \tag{2.9}$$

Since $\underline{\underline{A}}^{-1} = (\det \underline{\underline{A}})^{-1}(\underline{\underline{A}}^{\text{cof}})^T$, we see immediately that (2.9) is equivalent to (2.6), and that both definitions given in (2.2) and (2.3) yield the same result.

Example 2: $\epsilon = \hat{\epsilon}(\underline{\underline{A}}) = \operatorname{tr}(\underline{\underline{A}}^m)$, m being any positive integer.

Since we have yet to introduce the notion of the gradient of a matrix-valued function, we must rule out the possibility of calculating the gradient of $\hat{\epsilon}$ using the chain rule. To apply the definition of the gradient of $\hat{\epsilon}$ as given in (2.2), it is necessary to develop an expansion of the function $\hat{\epsilon}$ in terms of the components of A. We observe that for arbitrary positive integer m, the expansion of the matrix A^m is cumbersome, and it is not practical to find the gradient of $\hat{\epsilon}$ using (2.2).

However, the definition given in (2.3) does lead us to an answer:

$$\begin{aligned}\operatorname{tr}\{\underline{\underline{\epsilon_A}}^T \underline{\underline{C}}\} &= \frac{d}{ds}\operatorname{tr}\{(\underline{\underline{A}} + s\underline{\underline{C}})^m\}_{s=0} \\ &= \frac{d}{ds}\ \{(\underline{\underline{A}}^m + (\underline{\underline{A}}^{m-1}\underline{\underline{C}} + \underline{\underline{A}}^{m-2}\underline{\underline{C}}\underline{\underline{A}} + \ . \ . \ . + \underline{\underline{C}}\underline{\underline{A}}^{m-1})s + \ . \ . \ .)\}_{s=0} \\ &= \operatorname{tr}\{\underline{\underline{A}}^{m-1}\underline{\underline{C}} + \underline{\underline{A}}^{m-2}\underline{\underline{C}}\underline{\underline{A}} + \ . \ . \ . + \underline{\underline{C}}\underline{\underline{A}}^{m-1}\} \\ &= \operatorname{tr}\{m\underline{\underline{A}}^{m-1}\underline{\underline{C}}\},\end{aligned} \tag{2.10}$$

[3] For a rigorous exposition of the notion of a principal invariant of a matrix or a second order tensor, see Ericksen [2, p. 832].

since the trace operator is linear and $\mathrm{tr}(\underline{\underline{A}}\underline{\underline{B}})=\mathrm{tr}(\underline{\underline{B}}\underline{\underline{A}})$. Since $\underline{\underline{C}}$ *is arbitrary, we conclude that* $\underline{\underline{\epsilon_A}}=m(\underline{\underline{A}}^{m-1})^T$.

EXAMPLE 3: $\epsilon=\hat{\epsilon}(\underline{\underline{A}})=\det(\underline{\underline{A}}^2+\underline{\underline{B}}),\underline{\underline{B}}$ being a constant matrix.

In this case, both definitions given in (2.2) and (2.3) are not practical for us to evaluate the gradient of the scalar function $\hat{\epsilon}$. The only reasonable alternative is to use the chain rule in conjunction with a practical way of evaluating the gradient of a matrix-valued function as to be presented in the next section.

3. Gradient of a Matrix-Valued Function of a Matrix Argument

Let $f\equiv\hat{f}(A)$ define a matrix-valued function $\hat{f}$ of a matrix argument A where both f and A are square matrices of order n with components f_{km} and A_{rs} respectively, and the n^2 component functions $\hat{f}_{km}$ of $\hat{f}$ are defined as follows:

$$f_{km}\equiv\hat{f}_{km}(A_{rs})\equiv\hat{f}_{km}(\underline{\underline{A}}). \tag{3.1}$$

If each component function $\hat{f}_{km}$ is differentiable, the set of the first partial derivatives of $\hat{f}_{km}$, i.e., $\{D_{pq}\hat{f}_{km}\}$, can be defined as the gradient of the function $\hat{f}$ to be denoted by $\nabla\hat{f}$. To emphasize the need for four indices to specify f_A which stands for the value of $\nabla\hat{f}$ for a given A, we introduce the unusual four-bar notation as it appears in the following definition:

$$\underline{\underline{\underline{\underline{f_A}}}}\equiv[(f_A)_{kmpq}]\equiv[D_{pq}\hat{f}_{km}(A_{rs})]=[(\nabla\hat{f})_{kmpq}(A_{rs})]\equiv\underline{\underline{\underline{\underline{\nabla\hat{f}}}}}(\underline{\underline{A}}). \tag{3.2}$$

Clearly f_A is not a square matrix in the usual sense, and, therefore, is not suitable for calculations in matrix notation. Following [1], we introduce the so-called contraction operation on f_A with respect to an arbitrary square matrix C whose order is the same as that of A:

$$\underline{\underline{f_A[C]}}\equiv\left[\sum_{p=1}^{n}\sum_{q=1}^{n}(f_A)_{kmpq}C_{pq}\right]. \tag{3.3}$$

The new quantity, $f_A[C]$, to be known as the gradient of $\hat{f}$ with respect to A and contracted with C, requires only two indices for component representation. Hence the symbol $f_A[C]$ will replace f_A wherever matrix operations are used.

The definition of the gradient of $\hat{f}$ as given in (3.2) is equivalent to the following alternative definition based on the chain rule:

$$\underline{\underline{f_A[C]}}=\frac{d}{ds}\underline{\underline{\hat{f}}}(\underline{\underline{A}}+s\underline{\underline{C}}), \tag{3.4a}$$

or, in component notation,

$$\sum_{p=1}^{n}\sum_{q=1}^{n}(f_A)_{kmpq}C_{pq}=\frac{d}{ds}\hat{f}_{km}(A_{rt}+sC_{rt}). \tag{3.4b}$$

As a rule, both definitions given in (3.2) and (3.4) are useful for simple matrix-valued functions such as those listed below:

$$\underline{\underline{\hat{f}}}(\underline{\underline{A}})=\underline{\underline{A}}; \qquad \underline{\underline{f_A[C]}}=\underline{\underline{C}}. \tag{3.5}$$

$$\hat{f}(\underline{\underline{A}}) = \underline{\underline{A}}^T; \qquad \underline{\underline{f_A[C]}} = \underline{\underline{C}}^T. \tag{3.6}$$

$$\underline{\underline{\hat{f}(A)}} = \underline{\underline{BA}},\ B \text{ being a constant matrix;}\ \underline{\underline{f_A[C]}} = \underline{\underline{BC}}. \tag{3.7}$$

$$\underline{\underline{\hat{f}(A)}} = \underline{\underline{AB}},\ B \text{ being a constant matrix,}\ \underline{\underline{f_A[C]}} = \underline{\underline{CB}}. \tag{3.8}$$

For moderately complicated matrix-valued functions such as $\hat{f}(A) = A^m$, m being a positive integer greater than 1, the matrix definition given by (3.4a) is far superior and sometimes becomes the sole means of evaluating the gradient of a matrix-valued function. The reader can easily verify, using (3.4a), the following useful result: (Note: $\underline{\underline{A}}^\circ \equiv \underline{\underline{1}}$).

$$\underline{\underline{\hat{f}}}(\underline{\underline{A}}) = \underline{\underline{A}}^m,\ m = 2,3,4,\ \ldots\ ; \qquad \underline{\underline{f_A[C]}} = \sum_{i=0}^{m-1} \underline{\underline{A}}^i \underline{\underline{C}} \underline{\underline{A}}^{m-i-1}. \tag{3.9}$$

4. Product Rule for Differentiation of Matrix-valued Functions

Let $\hat{f}$ be the product of two matrix-valued functions $\hat{g}$ and $\hat{h}$ with $f \equiv \hat{f}(A) \equiv \hat{g}(A)\hat{h}(A) \equiv gh$, where the order of the matrix multiplication is important. The product rule for partial differentiation yields the gradient of $\hat{f}$ in the following matrix notation:

$$\underline{\underline{f_A[C]}} = \underline{\underline{g_A[C]}}\,\underline{\underline{h}} + \underline{\underline{g}}\,\underline{\underline{h_A[C]}} \qquad \text{for} \qquad \underline{\underline{f}} = \underline{\underline{g}}\,\underline{\underline{h}} \tag{4.1}$$

Using the elementary formulas given in (3.5) and (3.6), we obtain immediately the following formula based on (4.1):

$$\underline{\underline{\hat{f}}}(\underline{\underline{A}}) = \underline{\underline{A}}^T\underline{\underline{A}}; \qquad \underline{\underline{f_A[C]}} = \underline{\underline{C}}^T\underline{\underline{A}} + \underline{\underline{A}}^T\underline{\underline{C}}. \tag{4.2}$$

To derive the formula for the gradient of the matrix inversion operator, we apply the product rule to the identity $\underline{\underline{A}}^{-1}\underline{\underline{A}} = \underline{\underline{1}}$:

$$\underline{\underline{\hat{f}}}(\underline{\underline{A}}) = \underline{\underline{A}}^{-1}; \qquad \underline{\underline{f_A[C]}}\,\underline{\underline{A}} + \underline{\underline{A}}^{-1}\underline{\underline{C}} = \underline{\underline{O}},\ \text{i.e.,} \qquad \underline{\underline{f_A[C]}} = -\underline{\underline{A}}^{-1}\underline{\underline{C}}\,\underline{\underline{A}}^{-1}. \tag{4.3}$$

Using the product rule and (4.3), the reader can easily verify by induction:

$$\underline{\underline{\hat{f}}}(\underline{\underline{A}}) = \underline{\underline{A}}^{-m},\ m = 2,3,\ldots\ ; \qquad \underline{\underline{f_A[C]}} = -\sum_{i=0}^{m-1} \underline{\underline{A}}^{-m+i}\,\underline{\underline{C}}\,\underline{\underline{A}}^{-1+i}. \tag{4.4}$$

Whenever the inverse of a matrix is mentioned, the restriction to the class of square matrices with nonzero determinants will be understood.

5. Chain Rule for Differentiation of Scalar- and Matrix-Valued Functions

Consider a scalar-valued function $\hat{\epsilon}$ of a matrix argument $\underline{\underline{A}}$ whose components A_{km} are functions of a single scalar parameter t. The chain rule for differentiation with respect to t assumes the following form in component notation:

$$\eta(t) \equiv \hat{\epsilon}(A_{km}(t)); \qquad \dot{\eta}(t) = \sum_{k=1}^{n}\sum_{m=1}^{n} (\epsilon_A)_{km}\dot{A}_{km}(t), \tag{5.1a}$$

where the dot symbol denotes the operator $\frac{d}{dt}$. In matrix notation, (5.1a) becomes

$$\eta(t) \equiv \hat{\epsilon}(\underline{\underline{A}}(t)); \qquad \dot{\eta}(t) = \mathrm{tr}(\underline{\underline{\epsilon_A^T}}\, \underline{\underline{\dot{A}}}(t)). \tag{5.1b}$$

In applications, it is common to work with a scalar-valued function $\hat{\phi}$ of a matrix argument f which depends on another matrix argument A. The chain rule for differentiation with respect to A can be written in the following component notation:

$$\hat{\epsilon}(A_{rt}) \equiv \hat{\phi}(\hat{f}_{uv}(A_{ij})); \qquad (\epsilon_A)_{pq} = \sum_{k=1}^{n} \sum_{m=1}^{n} (\phi_f)_{km} (f_A)_{kmpq}. \tag{5.2a}$$

To write (5.2a) in matrix notation, let us contract both sides of (5.2a) with an arbitrary matrix C;

$$\hat{\epsilon}(\underline{\underline{A}}) \equiv \underline{\underline{\hat{\phi}}}\,(\hat{f}\,(A)); \qquad \mathrm{tr}(\underline{\underline{\epsilon_A^T}}\;\underline{\underline{C}}) = \mathrm{tr}(\underline{\underline{\phi_f^T\, f_A[C]}}). \tag{5.2b}$$

Returning to Example 3 given in section 2, we are now equipped to evaluate the gradient of the function $\hat{\epsilon}$ defined by $\epsilon = \hat{\epsilon}(\underline{\underline{A}}) = \det(\underline{\underline{A}}^2 + \underline{\underline{B}})$, B being a constant matrix. Using (2.9), (3.9) and the chain rule given by (5.2b), we have:

$$\mathrm{tr}\{\underline{\underline{\epsilon_A^T}}\,\underline{\underline{C}}\} = \mathrm{tr}\,\{\det(\underline{\underline{A}}^2 + \underline{\underline{B}})(\underline{\underline{A}}^2 + \underline{\underline{B}})^{-1}(\underline{\underline{A}}\,\underline{\underline{C}} + \underline{\underline{C}}\,\underline{\underline{A}})\}.$$

Since $\underline{\underline{C}}$ is arbitrary, we have $\underline{\underline{\epsilon_A^T}} = \det(\underline{\underline{A}}^2 + \underline{\underline{B}})\,[(\underline{\underline{A}}^2 + \underline{\underline{B}})^{-1}\underline{\underline{A}} + \underline{\underline{A}}(\underline{\underline{A}}^2 + \underline{\underline{B}})^{-1}]$.

6. A Collection of Some Elementary Formulas in "Matrix Calculus"

Based on the product rule and the chain rule for differentiation in matrix notation as presented in the last two sections, a calculus of differentiable functions of square matrices, to be referred

Formula number	Function	Gradient	Remark
1	$f(A) = A$	$f_A[C] = C$	
2	$\hat{f}(A) = A^T$	C^T	
3	$\hat{f}(A) = B\,A$	$B\,C$	B being a constant matrix.
4	$\hat{f}(A) = A\,B$	$C\,B$	B being a constant matrix.
5	$\hat{f}\,(A) = A^M$	$\sum_{i=0}^{m-1} A^i\, C\, A^{m-i-1}$	$m = 2, 3, 4, \ldots$
6	$\hat{f}(A) = A^T A$	$C^T A + A^T C$	
7	$\hat{f}(A) = A^{-1}$	$-A^{-1}\, C\, A^{-1}$	
8	$\hat{f}(A) = A^{-m}$	$-\sum_{i=0}^{m=1} A^{-m+i}\, C\, A^{-1+i}$	$m = 2, 3, 4, \ldots$
9	$\hat{\epsilon}(A) = \mathrm{tr}\, A$	$\epsilon_A = \quad 1$	First invariant.
10	$\hat{\epsilon}(A) = \det A$	$\det A\; (A^{-1})^T$	nth invariant.
11	$\hat{\epsilon}(A) = \phi(A^T)$	$(\phi_{A^T})^T$	
12	$\hat{\epsilon}(A) = \hat{\phi}(B\,A\,D)$	$B^T\,(\phi_{BAD})\,D^T$	B, D being constant matrices.
13	$\hat{\epsilon}(A) = \hat{\phi}(A^{-1})$	$-(A^{-1})^T\,(\phi_{A^{-1}})\,(A^{-1})^T$	
14	$\hat{\epsilon}(A) = \hat{\phi}(A\,A^T)$	$2\,(\phi_{AA^T})\,A$	Note $(\phi_{AA^T})^T = (\phi_{AA^T})$.
15	$\hat{\epsilon}(A) = \hat{\phi}(A^T A)$	$2\,A\,(\phi_{A^T A})$	Note $(\phi_{A^T A})^T = (\phi_{A^T A})$.
16	$\hat{\epsilon}(A) = \mathrm{tr}\, A^{-1} \det A$	$\det A\; (A^{-1})^T\,(1\,\mathrm{tr}\, A^{-1} - (A^{-1})^T)$	Second invariant for $n = 3$.

to as the "matrix calculus," can be defined in an analogous way as the elementary theory of calculus based on the field of real or complex numbers. Obviously, for matrix calculus, the underlying mathematical object is not a field, but a noncommutative ring, i.e., the ring of square matrices of order n over the familiar ring of differentiable functions. An excellent account of the theory of matrices over rings was given recently by Newman [3], but here we merely present a collection of some elementary formulas in "matrix calculus" without studying its mathematical structure. For the convenience of the reader, the table on page 102 lists some of the most commonly used formulas in matrix calculus.

Using the properties of the operator "trace" as listed in section 1, we observe that the derivation for formulas Nos. 11–15 presents no difficulty. For example, formula No. 13 can be derived as follows:
From (5.2b) and formula 7, we have

$$\mathrm{tr}\,(\underline{\underline{\epsilon_A^T}}\,\underline{\underline{C}}) = \mathrm{tr}\,(-(\underline{\underline{\phi_{A^{-1}}}})^T \underline{\underline{A}}^{-1}\underline{\underline{C}}\underline{\underline{A}}^{-1}) = \mathrm{tr}\,(-\underline{\underline{A}}^{-1}(\underline{\underline{\phi_{A^{-1}}}})^T \underline{\underline{A}}^{-1}\underline{\underline{C}}). \tag{6.1}$$

Since C is arbitrary, we obtain immediately the desired result.

7. Applications

To illustrate the ease with which certain problems in continuum physics can be treated by using some of the formulas listed in the last section, we shall present two examples in continuum mechanics:

EXAMPLE 1: (All indices i, j, k, m, p, etc. range from 1 to 3.)

Let the material coordinates of a particle in a continuous body be denoted by X^i. Let the position coordinates of the same particle at time t be given by $x^k \equiv \hat{x}^k(X^i, t)$. Two basic quantities can be defined:

$$\underline{\underline{F}} \equiv \hat{F}^k_i(X^j, t) \equiv \frac{\partial \hat{x}^k}{\partial X^i}(X^j, t); \qquad \text{(deformation gradient);} \tag{7.1}$$

$$\underline{v} \equiv \hat{v}^k(X^j, t) \equiv \frac{\partial \hat{x}^k}{\partial t}(X^j, t); \qquad \text{(velocity vector).} \tag{7.2}$$

It is useful to express X^i as functions of x^k and t so that the velocity components have the alternative representation $v^k = \tilde{v}^k(x^m, t)$. This allows us to define another useful quantity:

$$\underline{\underline{L}} \equiv \hat{L}^k_m(x^p, t) \equiv \frac{\partial \tilde{v}^k}{\partial x^m}(x^p, t), \qquad \text{(velocity gradient).} \tag{7.3}$$

An important relation to be needed later follows immediately from the above definitions and the interchange of the order of partial differentiations:

$$\dot{\underline{\underline{F}}} = \underline{\underline{L}}^T \underline{\underline{F}}, \tag{7.4}$$

where the dot symbol denotes the partial derivative with respect to t holding the material coordinates X^i constant.

The notion of "mass" of a continuous body leads to two notions of "mass density", namely, the mass density ρ_R with respect to a unit volume in the reference configuration where each particle is labeled with the material coordinates X^i, and the mass density ρ_t with respect to a unit volume in the spatial configuration at time t where each particle is observed to occupy the position at coordinates x^k. The two mass densities are of course related:

$$\rho_R = \rho_t \det \underline{\underline{F}}. \tag{7.5}$$

The law of the conservation of mass states that $\dot{\rho}_R = 0$. Using (5.1b), formula No. 10, and the relation (7.4), we obtain the well known "equation of continuity" in classical mechanics: (Note: $\operatorname{div} \underline{v} \equiv \sum_{k=1}^{3} \frac{\partial v^k}{\partial x^k}$.)

$$\begin{aligned}
\dot{\rho}_R &= \dot{\rho}_t \det \underline{\underline{F}} + \rho_t \overline{(\det \underline{\underline{F}})} \\
&= \dot{\rho}_t \det \underline{\underline{F}} + \rho_t \operatorname{tr} (\det \underline{\underline{F}}\, (\underline{\underline{F}}^{-1}) \dot{\underline{\underline{F}}}) \\
&= \dot{\rho}_t \det \underline{\underline{F}} + \rho_t \det \underline{\underline{F}} \operatorname{tr} (\underline{\underline{F}}^{-1} \underline{\underline{L}}^T \underline{\underline{F}}) \\
&= \dot{\rho}_t \det \underline{\underline{F}} + \rho_t \det \underline{\underline{F}} \operatorname{tr} (\underline{\underline{L}}^T) \\
&= \det \underline{\underline{F}}\, (\dot{\rho}_t + \rho_t \operatorname{div} \underline{v})
\end{aligned}$$

Since $\det \underline{\underline{F}} \neq 0$, $\dot{\rho}_R = 0$ implies $\dot{\rho}_t + \rho_t \operatorname{div} \underline{v} = 0$.

EXAMPLE 2:

One of the principles generally associated with the correct formulation of the constitutive equation of a material is known as the "Principle of Material Indifference" which means physically that the response of a material is independent of the observer. Let us confine our attention to "simple fluids" in the sense of Truesdell and Noll [1], where the most general constitutive equation may be written in the following form:

$$\underline{\underline{T}}(t) = \underline{\underline{\mathscr{G}}}[\underline{\underline{C}}_t(\tau);\, \rho_t]. \tag{7.6}$$

Here $T(t)$ is the Cauchy stress at time t, $C_t(\tau)$ is the relative right Cauchy-Green tensor defined by $(C_t(\tau))_{km} \equiv \sum_{p=1}^{3} \frac{\partial \hat{x}^p(\tau)}{\partial x^k(t)} \frac{\partial \hat{x}^p(\tau)}{\partial x^m(t)}$, ρ_t is the mass density at time t, and $\mathscr{G}$ is a functional of the history of $C_t(\tau)$, $-\infty < \tau \leq t$, with a parametric dependence on ρ_t. The principle of material indifference requires that the functional $\mathscr{G}$ satisfies the following relation for an arbitary orthogonal matrix Q:

$$\underline{\underline{\mathscr{G}}}[\underline{\underline{Q}}\underline{\underline{C}}_t(\tau)\underline{\underline{Q}}^T;\, \rho_t] = \underline{\underline{Q}}\underline{\underline{\mathscr{G}}}[\underline{\underline{C}}_t(\tau);\, \rho_t]\underline{\underline{Q}}^T. \tag{7.7}$$

Consider now the following constitutive equation of a simple fluid: [4]

$$\underline{\underline{T}}(t) = -\hat{p}(\rho_t)\underline{\underline{1}} + \rho_t \int_{-\infty}^{t} \underline{\underline{C}}_t(\tau)\, \frac{\partial \hat{U}}{\partial \underline{\underline{C}}_t(\tau)}\, d\tau, \tag{7.8}$$

[4] The theory of such a fluid was proposed by Bernstein, Kearsley and Zapas [4]. Additional results on the same theory including the derivation of equation (7.8) appeared in a recent manuscript by Fong and Simmons [5].

where $\hat{p}$ is a scalar-valued function of the mass density ρ_t, and $\hat{U}$ is a scalar-valued function of one matrix argument $\underline{\underline{C}}_t(\tau)$ and one scalar variable $t-\tau$, i.e., $U \equiv \hat{U}(\underline{\underline{C}}_t(\tau);\; t-\tau)$. Furthermore, $\hat{U}$ is required to satisfy the following condition for an arbitrary orthogonal Q:

$$\hat{U}(\underline{\underline{Q}}\underline{\underline{C}}_t(\tau)\underline{\underline{Q}}^T;\; t-\tau) = \hat{U}(\underline{\underline{C}}_t(\tau)\,;\; t-\tau). \tag{7.9}$$

Differentiating eq (7.9) with respect to $\underline{\underline{C}}_t(\tau)$ and applying the formula No. 12 as listed in the last section, we obtain

$$\underline{\underline{Q}}^T\underline{\underline{\hat{U}}}_C(\underline{\underline{Q}}\underline{\underline{C}}_t(\tau)\underline{\underline{Q}}^T;\; t-\tau)\underline{\underline{Q}} = \underline{\underline{\hat{U}}}_C(\underline{\underline{C}}_t(\tau);\; t-\tau), \tag{7.10}$$

where $\hat{U}_C$ denotes the partial gradient of $\hat{U}$ with respect to $C_t(\tau)$. We are now ready to show that eq (7.8), indeed, satisfies the principle of material indifference. Using (7.6) and (7.8), we calculate the left-hand side of (7.7):

$$\underline{\underline{\mathscr{G}}}\,[\underline{\underline{Q}}\underline{\underline{C}}_t(\tau)\underline{\underline{Q}}^T;\; \rho_t] = -\hat{p}(\rho_t)\underline{\underline{1}} + \rho_t \int_{-\infty}^{t} \underline{\underline{Q}}\underline{\underline{C}}_t(\tau)\underline{\underline{Q}}^T\underline{\underline{\hat{U}}}_C(\underline{\underline{Q}}\underline{\underline{C}}_t(\tau)\underline{\underline{Q}}^T;\; t-\tau)d\tau.$$

Substituting (7.10) into the above equation, we obtain

$$\underline{\underline{\mathscr{G}}}\,[\underline{\underline{Q}}\underline{\underline{C}}_t(\tau)\underline{\underline{Q}}^T;\; \rho_t] = -\hat{p}(\rho_t)\underline{\underline{1}} + \rho_t \int_{-\infty}^{t} \underline{\underline{Q}}\underline{\underline{C}}_t(\tau)\underline{\underline{Q}}^T\underline{\underline{Q}}\underline{\underline{U}}_C(\underline{\underline{C}}_t(\tau);\, t-\tau)\,\underline{\underline{Q}}^T d\tau.$$

$$= \underline{\underline{Q}}\left\{-\hat{p}(\rho_t)\underline{\underline{1}} + \rho_t \int_{-\infty}^{t} \underline{\underline{C}}_t(\tau)\underline{\underline{\hat{U}}}_C(\underline{\underline{C}}_t(\tau);\; t-\tau)\,d\tau\right\}\underline{\underline{Q}}^T$$

$$= \underline{\underline{Q}}\,\underline{\underline{\mathscr{G}}}\,[C_t(\tau)\,;\; \rho_t]\underline{\underline{Q}}^T = \text{R.H.S. of (7.7).} \qquad \textit{Q.E.D.}$$

It is clear from the above two examples that the advantage of adopting the matrix notation and applying formulas in "matrix calculus" lies mainly in the elegance in which higher-dimensional problems in continuum physics can be formulated. It is also clear that even though our list of formulas was prepared for functions of one matrix argument, their applications can be easily extended to functions of several matrix arguments.

I wish to thank Dick Kraft, John A. Simmons, Seldon L. Stewart and Justin C. Walker for many helpful comments.

8. References

[1] Truesdell, C., and Noll, W., The Non-Linear Field Theories of Mechanics, Encyclopedia of Physics, **III/3**, Ed., S. Flügge (Springer-Verlag, 1965).
[2] Ericksen, J. L., Tensor Fields, An Appendix in Encyclopedia pf Physics, **III/1**, Ed., S. Flügge (Springer-Verlag, 1960).
[3] Newman, M., Integral Matrices (Academic Press (in press)).
[4] Bernstein, B., Kearsley, E. A., and Zapas, L. J., A Study of stress relaxation with finite strain, Transactions of Society of Rheology, **7**, 391 (1963).
[5] Fong, J. T., and Simmons, J. A., The scalar potential of the BKZ theory for a compressible fluid (to be published).

(Paper 75B3&4–349)