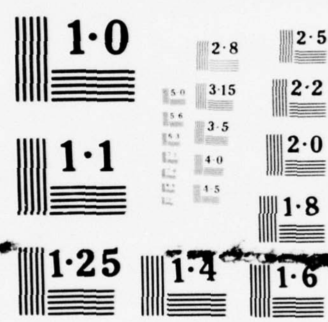
AD-A044 808   NATIONAL RANGE OPERATIONS DIRECTORATE WHITE SANDS MIS--ETC   F/G 12/2
NEW GRADIENT TECHNIQUES FOR TRACES OF FUNCTIONS OF RECTANGULAR --ETC(U)
SEP 77   J S PAPPAS, O N DALTON

UNCLASSIFIED   ACD-61-77   NL

1 OF 1
ADA
044808

END
DATE
FILMED
10-77
DDC

END
DATE
FILMED
10-77
DDC



NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

AD

(12)

NEW GRADIENT TECHNIQUES FOR TRACES OF FUNCTIONS OF RECTANGULAR MATRICES AND THEIR PSEUDO-INVERSES

BY

JAMES S. PAPPAS and OREN N. DALTON

SEPTEMBER 1977

ACD-61-77

MATHEMATICAL SERVICES BRANCH
ANALYSIS & COMPUTATION DIVISION
NATIONAL RANGE OPERATIONS DIRECTORATE
WHITE SANDS MISSILE RANGE, NEW MEXICO

Approved for public release; distribution unlimited.

407706

Destroy this report when no longer needed. Do not return to the originator.

Disclaimer

The findings of this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

ABSTRACT

The generalized inverse is of increasing importance for estimation and optimization in modern systems theory, because it both simplifies many problems and reveals underlying structures of theoretical importance. Optimization, using the gradient of the trace of products of matrix valued functions, including the generalized inverse, are presented in a novel state-space setting. A number of functions of such products of matrices, including general formulas, are derived for the first time to our knowledge. Tables of a number of them are given with some applications in the fields of estimation and optimization.

TABLE OF CONTENTS

NEW GRADIENT TECHNIQUES FOR TRACES OF FUNCTIONS OF RECTANGULAR MATRICES AND THEIR PSEUDO-INVERSES

I. INTRODUCTION. The state space approach to modern estimation and control theory requires optimal feedback or weighting matrices which minimize the trace of quadratic matrix functions. In the estimation problem the quadratic matrix function is the variance of the state estimation. For example, in the Kalman theory the best estimate requires that the error matrix, P, be minimized with respect to a weighting matrix W, where P propagates via the matrix Riccati equation.

$$\dot{P} = AP + PA^T - WHP - PH^TW^T + WRW^T$$

The novel vector and matrix partitioning techniques developed in this paper, eliminate much of the tedium of the analysis based on scalar sums and products. In addition, they reveal a depth of structure not otherwise apparent.

Section II is devoted to deriving gradient relations. Initially, considerable detail is exhibited to show necessary relationships; afterwards, this detail is bypassed. A table of gradients for a number of common important forms, principally linear and quadratic forms, is included at the end. Athans [1] presents some linear gradient forms for the trace as well as some for determinants, logs and exponentials; Tracy and Dwyer [2] present additional gradient forms some of which involve the Kronecker matrix product; and Neudecker [3] presents gradients involving traces, latent roots and determinants which are related to frequently occurring statistical econometric problems.

In Section III a general formula is developed for functions expressable as a power series. Particular reference is made to trigonometric and hyperbolic functions.

In Section IV applications for estimation and filtering are demonstrated by developing criteria for the continuous Kalman filter.

II. GRADIENT DERIVATIONS. Several useful relations involving the gradient of some scalar valued functions with respect to vectors and matrices are developed in this section. The cases are restricted to linear, quadratic and "cubic" matrix terms; the latter implying the generalized inverse. Throughout the derivations, use will be made of well known results of traces of matrices, namely

$$tr(A+B) = trA + trB$$

$$\text{tr } M = \text{tr } M^T \quad (1)$$

and the cyclic property of products. For example

$$\text{tr } \underset{p\times m}{R}\ \underset{m\times \ell}{S}\ \underset{\ell\times p}{T} = \text{tr } \underset{\ell\times p}{T}\ \underset{p\times m}{R}\ \underset{m\times \ell}{S} = \text{tr } \underset{m\times \ell}{S}\ \underset{\ell\times p}{T}\ \underset{p\times m}{R} \quad (2)$$

A. <u>LINEAR CASES</u>. Consider the $\ell\times\ell$ matrix function of the rectangular matrix X of size $p\times m$, that is

$$\underset{\ell\times\ell}{L_1} = \underset{\ell\times p}{A}\ \underset{p\times m}{X}\ \underset{m\times \ell}{B} \quad (3)$$

with trace

$$\text{tr } L_1 = \ell = \text{tr } AXB \quad \text{(Note this } \ell \text{ is not the matrix size.)} \quad (4)$$

the $m\times m$ matrix L_2

$$\underset{m\times m}{L_2} = \underset{m\times \ell}{B}\ \underset{\ell\times p}{A}\ \underset{p\times m}{X} \quad (5)$$

with trace

$$\text{tr } L_2 = \text{tr } L_1 = \ell \quad (6)$$

and the $p\times p$ matrix

$$\underset{p\times p}{L_3} = \underset{p\times m}{X}\ \underset{m\times \ell}{B}\ \underset{\ell\times p}{A} \quad (7)$$

and by Equation (2)

$$\ell = \text{tr } AXB = \text{tr } BAX = \text{tr } XBA \quad (8)$$

The differential of the trace is

$$d\ell = d \text{ tr } (AXB) = d \text{ tr } (BAX) = d \text{ tr } (XBA) \quad (9)$$

or

$$d\ell = \text{tr } \underset{\ell\times p}{A}\ \underset{p\times m}{dX}\ \underset{m\times \ell}{B} = \text{tr } \underset{m\times m}{BAdX}$$

$$= \text{tr } \underset{p\times p}{(dXBA)} \quad (10)$$

The differential of the m×m matrix can be written as

$$\underset{m\times m}{dL_2} = \underset{m\times p}{BA}\ \underset{p\times m}{dX} = \underset{m\times p}{\frac{\partial \ell}{\partial X^T}}\ \underset{p\times m}{dX} \qquad (11)$$

and the differential of the p×p matrix can be written as

$$\underset{p\times p}{dL_3} = \underset{p\times m}{dX}\ \underset{m\times p}{BA} = \underset{p\times m}{dX}\ \underset{m\times p}{\frac{\partial \ell}{X^T}} \qquad (12)$$

The differential of the traces are

$$d\ell = tr\left[\underset{p\times m}{\frac{\partial \ell}{\partial X^T}}\ \underset{m\times p}{dX}\right] = tr\left[\underset{p\times m}{dX}\ \underset{m\times p}{\frac{\partial \ell}{\partial X^T}}\right] \qquad (13)$$

and by Equation (10) and Equation (13)

$$\frac{\partial \ell}{\partial X^T} = \underset{m\times \ell}{B}\ \underset{\ell\times p}{A} \qquad (14)$$

or

$$\frac{\partial}{\partial X^T}\, tr\ \underset{\ell\times \ell}{(AXB)} = \frac{\partial}{\partial X^T}\, tr\ \underset{m\times m}{(BAX)} = \frac{\partial}{\partial X^T}\, \underset{p\times p}{(XBA)} = \underset{m\times \ell}{B}\ \underset{\ell\times p}{A} \qquad (15)$$

For the special case when the L_i are scalers, i.e.: A and B are vectors

$$\frac{\partial}{\partial X^T}\, tr\ \langle p)\ a\ \underset{p\times m}{X}\ b\ (m\rangle = b(m\rangle\langle p)a \qquad (16)$$

$$\frac{\partial tr}{\partial X^T}\left[b(m\rangle\langle p)a\ \underset{p\times m}{X}\right] = b(m\rangle\langle p)a \qquad (17)$$

and

$$\frac{\partial}{\partial X^T}\,\mathrm{tr}\left[\underset{p\times m}{X}\, b(m\rangle\langle p)a\right] = b(m\rangle\langle p)a \tag{18}$$

By Equation (3) for $m=\ell$ and $B=Im$

$$\underset{m\times p}{\frac{\partial \mathrm{tr}}{\partial X^T}}\left(\underset{m\times p}{A}\;\underset{p\times m}{X}\right) = \underset{m\times p}{A} \tag{19}$$

By Equation (3) for $\ell=p$ and $A=I_p$

$$\underset{m\times p}{\frac{\partial \mathrm{tr}}{\partial X^T}}\left(\underset{p\times m}{X}\;\underset{m\times p}{B}\right) = \underset{m\times p}{B} \tag{20}$$

In the above equation if $m=1$

$$\underset{1\times p}{\frac{\partial \mathrm{tr}}{\partial X^T}}\left[x(p\rangle\langle p)b\right] = \langle p)b \tag{21}$$

Since the trace of the dyadic product is the inner product

$$\mathrm{tr}\;(x\rangle\langle b) = \langle bx\rangle = \ell \tag{22}$$

one has

$$\ell\,\langle\frac{\partial}{\partial x} = \langle\frac{\partial \ell}{\partial x} = \langle b\; x\rangle\langle\frac{\partial}{\partial x} = \langle b \tag{23}$$

where the matrix

$$x\rangle\langle\frac{\partial}{\partial x} = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^p \end{bmatrix}\left[\frac{\partial}{\partial x_1},\cdots\frac{\partial}{\partial x_p}\right] = I \tag{24}$$

when the coordinates are all independent.

One can verify all of the previous results via the tedious process of partitioning. For example consider Equation (20)

$$\underset{p\times p}{L} = \underset{p\times m}{X} \; \underset{m\times p}{B} \tag{25}$$

and partition X into its column vectors and B into its row vectors to obtain

$$XB = \left[x(p\rangle_1, \cdots x(p\rangle_m \right] \begin{bmatrix} {}^1\langle p)b \\ \vdots \\ {}^m\langle p)b \end{bmatrix} = \sum_{i=1}^{m} x(p\rangle_i \, {}^i\langle p)b \tag{26}$$

the trace of Equation (26) is

$$\text{tr } XB = \text{tr}\left(x\rangle_1 {}^1\langle b \right) + \text{tr}\left(x\rangle_2 {}^2\langle b \right) + \cdots + \text{tr}\left(x\rangle_m {}^m\langle b \right) \tag{27}$$

or

$$\ell = \text{tr } XB = {}_1\langle bx\rangle_1 + {}_2\langle bx\rangle_2 + \cdots + {}_m\langle bx\rangle_m \tag{28}$$

$$= \ell_1 + \ell_2 + \cdots + \ell_m \tag{29}$$

The differential of ℓ is

$$d\ell = d\ell_1 + d\ell_2 + \cdots + d\ell_m \tag{30}$$

where each of the ℓ_i is a function of the vector $x\rangle_i$ or by Equation (13)

$$d\ell_i = {}_i\left\langle \frac{\partial \ell_i}{\partial x} \, dx \right\rangle_i \tag{31}$$

Repackaging Equation (31)

$$d\ell = {}_1\left\langle \frac{\partial \ell_i}{\partial x} \, dx(p \right\rangle_1 + \cdots + {}_m\left\langle \frac{\partial \ell_m}{\partial x} \, dx(p \right\rangle_m \tag{32}$$

By Equation (31) in Equation (28)

$$ {}_i\left\langle \frac{\partial \ell}{\partial x} \right. = {}_i\left\langle \frac{\partial \ell_i}{\partial x} \right. \tag{33}$$

Consider the product of the matrix of gradient vectors with the matrix of dX vectors that is

$$\begin{bmatrix} {}_1\left\langle p)\frac{\partial \ell}{\partial x} \right. \\ {}_2\left\langle p)\frac{\partial \ell}{\partial x} \right. \\ \vdots \\ {}_m\left\langle p)\frac{\partial \ell}{\partial x} \right. \end{bmatrix} \begin{bmatrix} dx(p\rangle_1, \cdots dx(p\rangle_m \end{bmatrix} \tag{34}$$

$$= \begin{bmatrix} {}_1\left\langle \frac{\partial \ell}{\partial x}\, dx \right\rangle_1 & \cdots & {}_1\left\langle \frac{\partial \ell}{\partial x}\, dx \right\rangle_m \\ \vdots & & \vdots \\ {}_m\left\langle \frac{\partial \ell}{\partial x}\, dx \right\rangle_1 & \cdots & {}_m\left\langle p)\frac{\partial \ell}{\partial x}\, dx \right\rangle_m \end{bmatrix} \tag{35}$$

$$= \underset{m\times p}{\frac{\partial \ell}{\partial X^T}} \; \underset{p\times m}{dX} \tag{36}$$

Clearly

$$\mathrm{tr}\left[\frac{\partial \ell}{\partial X^T}\, dX\right] = d\ell \tag{37}$$

The other results can be shown by similar partitioning and can be found in detail in reference [4].

B. QUADRATIC CASES. Consider the $\ell\times\ell$ matrix product

$$\underset{\ell\times\ell}{Q_1} = \underset{\ell\times p}{A}\ \underset{p\times m}{X}\ \underset{m\times m}{C}\ \underset{m\times p}{X^T}\ \underset{p\times\ell}{B} \quad (38)$$

or

$$\underset{\ell\times\ell}{Q_1} = \left[\underset{\ell\times m}{AX}\right]\left[\underset{m\times\ell}{C\ X^T\ B}\right] \quad (39)$$

Form the new matrix of size m×m from Equation (39) cyclically permuted via parentheses

$$\underset{m\times m}{Q_2} = \left[\underset{m\times\ell}{C\ X^T\ B}\right]\ \underset{\ell\times m}{AX} \quad (40)$$

Form a third matrix of size p×p by cyclically permuting the A matrix of Equation (39)

$$\underset{p\times p}{Q_3} = \underset{p\times m}{X}\ \underset{m\times m}{C}\ \underset{m\times p}{X^T}\ \underset{p\times\ell}{B}\ \underset{\ell\times p}{A} \quad (41)$$

The differential of Equation (40) is

$$\underset{m\times m}{dQ_2} = C\ dX^TBAX + CX^TBAdX \quad (42)$$

The first matrix on the right of Equation (42) under the trace transposition and permutation rules becomes

$$tr\ (CdX^TBAX) = tr\ C^TX^TA^TB^TdX \quad (43)$$

Using Equation (43) in the trace of Equation (42) yields

$$tr\ \underset{m\times m}{dQ_2} = tr\ (C^TX^TA^TB^T+CX^TBA)\ dX \quad (44)$$

The gradient factors of dQ_2 can be taken to be

$$\underset{m\times m}{dQ_2} = \frac{\partial tr dQ_2}{\underset{m\times p}{\partial X^T}} \underset{p\times m}{dX} \tag{45}$$

and by Equation (45) and Equation (44)

$$\frac{\partial tr}{\partial X^T} (XCX^T BA) = C^T X^T A^T B^T + CX^T BA \tag{46}$$

By Equation (38), Equation (40) and Equation (43) the traces are all equal, that is

$$tr\ Q_1 = tr\ Q_2 = tr\ Q_3 \tag{47}$$

and the

$$d\ tr\ Q_i = tr\ dQ_i \qquad i = 1, 2, 3 \tag{48}$$

hence

$$\frac{\partial tr}{\underset{m\times p}{\partial X^T}} \left[\underset{\ell\times p}{A} \ \underset{p\times m}{X} \ \underset{m\times m}{C} \ \underset{m\times p}{X^T} \ \underset{p\times \ell}{B} \right] = C^T X^T A^T B^T + CX^T BA \tag{49}$$

and

$$\frac{\partial tr}{\partial X^T} \left[CX^T BAX \right] = C^T X^T A^T B^T + CX^T BA \tag{50}$$

If C is square and equal to I, we obtain by Equation (46), Equation (49) and Equation (50)

$$\frac{\partial tr}{\partial X^T} \left[\underset{p\times p}{XX^T} BA \right] = \frac{\partial tr}{\partial X^T} \left[AXX^T B \right]$$
$$= \frac{\partial tr}{\partial X^T} \left[X^T BAX \right] = X^T (A^T B^T + BA) \tag{51}$$

If

$$BA = I \tag{52}$$

then Equation (51) becomes

$$\frac{\partial tr}{\partial X^T}\left[X^T X\right] = \frac{\partial tr}{\partial X^T}\left[X X^T\right] = 2\,\underset{m\times p}{X^T} \tag{53}$$

For the special case when Equation (49) is a scalar, one obtains

$$\frac{\partial}{\partial X^T}\left[\langle a X C X^T b\rangle\right] = C^T X^T a\rangle\langle b + C X^T b\rangle\langle a \tag{54}$$

and for C=I

$$\frac{\partial}{\partial X^T}\left[\langle a X X^T b\rangle = \underset{m\times p}{X^T}\left[a\rangle\langle b + b\rangle\langle a\right] \tag{55}$$

For p=1 by Equation (51)

$$\frac{\partial}{\partial X^T}\langle x B A x\rangle = \langle x\left(a\rangle\langle b + b\rangle\langle a\right) \tag{56}$$

For C=I in Equation (54)

$$\frac{\partial tr}{\partial X^T}\left[X^T b\rangle\langle a X\right] = X^T\left(a\rangle\langle b + b\rangle\langle a\right) \tag{57}$$

As a final quadratic case consider the (full rank) grammian-matrix

$$X^T X = G \tag{58}$$

and its inverse

$$(X^T X)^{-1} = G^{-1} \tag{59}$$

where

$$G^{-1} G = I \tag{60}$$

The differential of Equation (60) is

$$dG^{-1}G + G^{-1}dG = 0 \tag{61}$$

or

$$dG^{-1} = -G^{-1}dGG^{-1} \tag{62}$$

$$= -G^{-1}[dX^{T}X+X^{T}dX]\ G^{-1} \tag{63}$$

or

$$dG^{-1} = -G^{-1}dX^{T}XG^{-1} - G^{-1}X^{T}dXG^{-1} \tag{64}$$

The generalized inverse, for the full-rank case, is defined as

$$\underset{m\times p}{X^{*}} = (X^{T}X)^{-1}\ X^{T} = G^{-1}X^{T} \tag{65}$$

and its transpose is

$$\underset{p\times m}{X^{*T}} = \underset{p\times m}{X}\ \underset{m\times m}{G^{-1}} \tag{66}$$

Using Equation (65) and Equation (66) in Equation (64)

$$dG^{-1} = -\ G^{-1}\ dX^{T}X^{*T} - X^{*}dXG^{-1} \tag{67}$$

The trace of Equation (67) is

$$\mathrm{tr}\ dG^{-1} = -\ \mathrm{tr}\ (dX^{T}X^{*T}G^{-1}) - \mathrm{tr}\ (G^{-1}X^{*}dX)$$

Using the trace transpose property on the first right hand side term of Equation (68)

$$\mathrm{tr}\ dG^{-1} = -\ \mathrm{tr}\ (G^{-1}X^{*}dX+G^{-1}X^{*}dX) \tag{69}$$

$$= -\ 2\ \mathrm{tr}(G^{-1}X^{*}dX)$$

or

$$\frac{\partial tr}{\partial X^T_{m\times p}} (X^T X)^{-1}_{m\times m} = -2\,(X^T X)^{-1} X^* \tag{71}$$

Using Equation (65) in Equation (71)

$$\frac{\partial tr}{\partial X^T} (X^T X)^{-1} = -2\,(X^T X)^{-2}_{m\times m}\, X^T_{m\times p} \tag{72}$$

C. "CUBIC" CASE. The following cases do not involve cubics but the matrix X and X^T appear three times in the generalized inverse relation (full rank)

$$X^*_{m\times p} = (X^T X)^{-1}_{m\times m}\, X^T_{m\times p} \tag{73}$$

Consider the linear form in X* that is

$$Q_{m\times m} = X^*_{m\times p}\, B_{p\times m} \tag{74}$$

and

$$dQ = dX^* B \tag{75}$$

$$dQ = d\,[(X^T X)^{-1} X^T]\, B$$

$$= d(X^T X)^{-1}\, X^T B + (X^T X)^{-1}\, dX^T B \tag{76}$$

by Equation (63) in Equation (76)

$$dQ = -\,G^{-1}\,[dX^T X + X^T dX]\, G^{-1} X^T B + G^{-1} dX^T B \tag{77}$$

or by Equation (73) in Equation (77)

$$dQ = -\,G^{-1}[dX^T X + X^T dX]X^*B + G^{-1}dX^T B \tag{78}$$

Using the relation for the projector

$$\underset{p\times m}{X}\ \underset{m\times p}{X^*} = \underset{p\times p}{P} = P^T = P^2 \tag{79}$$

in Equation (78)

$$dQ = -\,G^{-1}\,dX^T\,PB - X^*dXX^*B + G^{-1}\,dX^T B \tag{80}$$

or

$$dQ = G^{-1}\,dX^T\,(I-P)\,B - X^*\,dXX^*B \tag{81}$$

The orthogonal compliment projector is

$$\tilde{P} = I - P = \tilde{P}^T = \tilde{P}^2 \tag{82}$$

Using Equation (82) and Equation (81)

$$dQ = G^{-1}\,dX^T\tilde{P}B - X^*dXX^*B \tag{83}$$

The trace relations yield

$$\mathrm{tr}\,dQ = +\,\mathrm{tr}(dX^T\tilde{P}BG^{-1}) - \mathrm{tr}\,(X^*BX^*dX) \tag{84}$$

and transpose-wise

$$\mathrm{tr}\,dQ = \mathrm{tr}\,(\,G^{-1}B^T\tilde{P}dX) - \mathrm{tr}\,X^*BX^*dX \tag{85}$$

or

$$\mathrm{tr}\,dQ = \mathrm{tr}\,[(\,G^{-1}B^T\tilde{P} - X^*BX^*)dX] \tag{86}$$

Expressing dQ as

$$\underset{m\times m}{dQ} = \underset{m\times p}{\frac{\partial\,\mathrm{tr}dQ}{\partial X^T}}\ \underset{p\times m}{dX} \tag{87}$$

one obtains

$$\boxed{\frac{\partial tr}{\partial X^T}(X^*B)_{m\,m} = G^{-1} B^T\tilde{P} - X^* BX^*} \tag{88}$$

D. <u>POWERS OF $(X^*B)^n$.</u> The following cases are of higher degree than third. Consider n=2 then the Q of Equation (74) has square

$$Q^2 = X^*BX^*B \tag{89}$$

and

$$dQ^2 = dX^*BX^*B + X^*BdX^*B \tag{90}$$

By Equation (75)

$$dQ = dX^*B \tag{91}$$

hence Equation (91) in Equation (90)

$$dQ^2 = dQX^*B + X^*BdQ \tag{92}$$

By Equation (89) let

$$q_2 = tr\ Q^2 \tag{93}$$

then

$$dQ^2 = \frac{\partial q_2}{\partial X^T} dX = dQQ + QdQ \tag{94}$$

The trace of Equation (92) is

$$tr\ dQ^2 = tr\ 2X^*BdQ \tag{95}$$

By Equation (94) and Equation (95)

$$\frac{1}{2} \, tr \, dQ^2 = tr \, Q \, dQ = tr \, X^*BdQ \tag{96}$$

Using dQ given by Equation (83) in Equation (96)

$$tr \, X^*BdQ = tr \, \{X^*B \, (G^{-1}dX^T\tilde{P}B - X^*dXX^*B)\}$$

$$= tr \, X^*BG^{-1}dX^T\tilde{P}B - tr \, X^*BX^*dXX^*B \tag{97}$$

$$= tr \, dX^T \, (\tilde{P}BX^*BG^{-1}) - tr \, X^*BX^*BX^*dX \tag{98}$$

$$= tr \, (G^{-1}B^TX^{*T}B^T\tilde{P})dX - tr \, Q^2X^*dX \tag{99}$$

$$\frac{1}{2} \, tr \, dQ^2 = tr \, \{(G^{-1}B^TX^{*T}B^T\tilde{P} - Q^2X^*) \, dX\} \tag{100}$$

or

$$\frac{1}{2} \frac{\partial q_2}{\partial X^T} = G^{-1}Q^TB^T\tilde{P} - Q^2X^* \tag{101}$$

That is,

$$\frac{\partial q_2}{\partial X^T} = 2 \, (G^{-1}Q^TB^T\tilde{P} - Q^2X^*) \tag{102}$$

Let n=3 and

$$Q^3 = QQQ \tag{103}$$

with

$$tr \, Q^3 = q_3 \tag{104}$$

The differential of Equation (103) is

$$dQ^3 = dQQ^2 + QdQQ + Q^2dQ \tag{105}$$

$$= \frac{\partial q_3}{\partial X^T} \, dX \tag{106}$$

and the trace of Equation (105) is

$$\operatorname{tr} dQ^3 = 3 \operatorname{tr} Q^2 dQ \tag{107}$$

By Equation (83) in Equation (107)

$$\operatorname{tr} dQ^3 = 3 \operatorname{tr} Q^2 \, [G^{-1} dX^T \tilde{P} B - X^* dXQ] \tag{108}$$

Using the commuting and permuting properties of trace one obtains

$$\underset{m\times p}{\frac{\partial q_3}{\partial X^T}} = 3\left[G^{-1} \, (Q^T)^2 \, B^T \tilde{P} - Q^3 X^*\right] \tag{109}$$

For n=1,2, and 3 by Equation (88), Equation (101) and Equation (109),

$$\frac{\partial q_1}{\partial X^T} = G^{-1} B^T \tilde{P} - QX^* \tag{110}$$

$$\frac{\partial q_2}{\partial X^T} = 2\left[G^{-1} Q^T B^T \tilde{P} - Q^2 X^*\right] \tag{111}$$

$$\frac{\partial q_3}{\partial X^T} = 3\left[G^{-1} (Q^T)^2 \, B^T \tilde{P} - Q^3 X^*\right] \tag{112}$$

This gives the inductive step, therefore for

$$\underset{m\times m}{Q} = \underset{m\times p}{X^*} \; \underset{p\times m}{B} \tag{113}$$

$$\underset{m\times m}{G^{-1}} = \underset{m\times m}{(X^T X)^{-1}} \tag{114}$$

$$\underset{m\times p}{X^*} = G^{-1} X^T \tag{115}$$

$$\underset{p\times p}{P} = XX^* \tag{116}$$

$$\tilde{P} = I - P \tag{117}$$

$$q_n = \text{tr}\ Q^n; \quad n = 1,2,\cdots \tag{118}$$

we have

$$\frac{\partial q_n}{\partial X^T} = n\ [G^{-1}(Q^T)^{n-1}B^T\tilde{P} - Q^n X^*] \tag{119}$$

A number of special cases follow from Equation (119). Suppose X is square and full rank, then

$$\underset{m\times m}{X^*} = X^{-1} \tag{120}$$

and

$$Q = X^{-1}B \tag{121}$$

Then

$$P = XX^{-1} = I \tag{122}$$

and

$$\tilde{P} = 0 \tag{123}$$

hence

$$\frac{\partial\, \text{tr}}{\partial X^T}\ (X^{-1}B)^n = -\ n\ (X^{-1}B)^n\ X^{-1} \tag{124}$$

Suppose now that B=I then

$$\frac{\partial\, \text{tr}}{\partial X^T}\ (X^{-1})^n = -\ n\ (X^{-1})^n\ X^{-1} \tag{125}$$

or

$$\frac{\partial \, tr}{\partial X^T} (X^{-1})^n = nX^{-(n+1)} \tag{126}$$

E. <u>POWERS OF $(XB)^n$</u>. The powers of

$$\underset{p\times p}{Q^n} = (XB)^n \tag{127}$$

can be obtained in a similar manner. By Equation (20)

$$\frac{\partial q_1}{\partial X^T} = \underset{m\times p}{B} \tag{128}$$

Let

$$Q^2 = (XB)^2 \tag{129}$$

$$q_2 = tr\ Q^2 \tag{130}$$

and

$$dQ^2 = dQQ + QdQ \tag{131}$$

$$tr\ dQ^2 = 2\ tr\ (QdQ) \tag{132}$$

$$= 2\ tr\ QdXB = 2\ tr\ (BQdX) \tag{133}$$

or

$$\frac{\partial q_2}{\partial X^T} = 2\ BQ \tag{134}$$

Repeating the arguments as before one obtains

$$\frac{\partial}{\partial X^T} [tr\ (XB)^n] = n \underset{m\times p}{B} (\underset{p\times p}{X\ B})^{n-1} \tag{135}$$

For X square, full rank, and

$$B = I$$

Equation (135) becomes

$$\frac{\partial \mathrm{tr} X^n}{\partial X^T} = nX^{n-1} \tag{136}$$

F. TABLES OF GRADIENTS

1. Linear Forms:

$$\left.\begin{matrix} \dfrac{\partial \mathrm{tr}}{\partial X^T}\left(\underset{\ell\times p}{A} \; \underset{p\times m}{X} \; \underset{m\times \ell}{B} \right) \\[2ex] \dfrac{\partial \mathrm{tr}}{\partial X^T}\left(\underset{m\times \ell}{B} \; \underset{\ell\times p}{A} \; \underset{p\times m}{X} \right) \\[2ex] \dfrac{\partial \mathrm{tr}}{\partial X^T}\left(\underset{p\times m}{X} \; \underset{m\times \ell}{B} \; \underset{\ell\times p}{A} \right) \end{matrix}\right\} = BA$$

$$\left.\begin{matrix} \dfrac{\partial \mathrm{tr}}{\partial X^T} \langle aXb\rangle \\[2ex] \dfrac{\partial \mathrm{tr}}{\partial X^T} [b(m\rangle\langle p)a \; \underset{p\times m}{X}] \\[2ex] \dfrac{\partial \mathrm{tr}}{\partial X^T} [\underset{p\times m}{X} \; b(m\rangle\langle p)a] \end{matrix}\right\} = b(m\rangle\langle p)a$$

$$\frac{\partial \mathrm{tr}}{\partial X^T} [\underset{m\times p}{A} \; \underset{p\times m}{X}] = \underset{m\times p}{A}$$

$$\frac{\partial \mathrm{tr}}{\partial X^T} [\underset{p\times m}{X} \; \underset{m\times p}{B}] = \underset{m\times p}{B}$$

$$\frac{\partial \mathrm{tr}}{\partial X^T}\left(x(p\rangle\langle p)b \right) = \langle p)b$$

2. <u>Quadratic Forms:</u>

$$\left.\begin{array}{l}
\dfrac{\partial tr}{\partial X^T}\left[\underset{\ell\times p}{A}\;\underset{p\times m}{X}\;\underset{m\times m}{C}\;\underset{m\times p}{X^T}\;\underset{p\times \ell}{B}\right] \\[3ex]
\dfrac{\partial tr}{\partial X^T}\left[\underset{m\times m}{C}\;\underset{m\times p}{X^T}\;\underset{p\times \ell}{B}\;\underset{\ell\times p}{A}\;\underset{p\times m}{X}\right] \\[3ex]
\dfrac{\partial tr}{\partial X^T}\left[\underset{p\times m}{X}\;\underset{m\times m}{C}\;\underset{m\times p}{X^T}\;\underset{p\times \ell}{B}\;\underset{\ell\times p}{A}\right]
\end{array}\right\} = C^T X^T A^T B^T + C X^T B A$$

$$\left.\begin{array}{l}
\dfrac{\partial tr}{\partial X^T}\left[X X^T B A\right] \\[3ex]
\dfrac{\partial tr}{\partial X^T}\left[A X X^T B\right] \\[3ex]
\dfrac{\partial tr}{\partial X^T}\left[X^T B A X\right]
\end{array}\right\} = X^T (A^T B^T + BA) .$$

$$\frac{\partial tr}{\partial X^T}\left[X^T X\right] = 2X^T$$

$$\left.\begin{array}{l}
\dfrac{\partial}{\partial X^T}\left(\langle a X X^T b\rangle\right) \\[3ex]
\dfrac{\partial tr}{\partial X^T}\left(X^T b\rangle\langle a X\right)
\end{array}\right\} = X^T \left(a\rangle\langle b + b\rangle\langle a\right)$$

$$\frac{\partial}{\partial X^T}\left(\langle x B A x\rangle\right) = \langle x\left(a\rangle\langle b + b\rangle\langle a\right)$$

$$\frac{\partial tr}{\partial X^T}\left(X^T X\right)^{-1} = -\,2\left(X^T X\right)^{-2} X^T$$

3. "Cubic" Forms and Others: The generalized inverse

$$X^* = (X^T X)^{-1} X^T = G^{-1} X^T$$

$$\frac{\partial tr}{\partial X^T} (\underset{m\times p}{X^*} \underset{p\times m}{B}) = (X^T X)^{-1} B^T \tilde{P} - X^* B X^*$$

where the projectors are

$$\tilde{P} = I - P$$

$$P = XX^*$$

$$G = X^T X$$

$$\frac{\partial tr}{\partial X^T} (X^* B)^n = n \, [G^{-1} (Q^T)^{n-1} B^T \tilde{P} - Q^n X^*]$$

$$\frac{\partial tr}{\partial X^T} (X^{-1} B)^n = - n \, (X^{-1} B)^n X^{-1}$$

$$\frac{\partial tr}{\partial X^T} (X^{-1})^n = - n \, X^{-(n+1)}$$

$$\frac{\partial tr}{\partial X^T} (XB)^n = nB(XB)^{n-1}$$

$$\frac{\partial tr}{\partial X^T} (X^n) = nX^{n-1}$$

III. SOME FUNCTIONS OF GENERALIZED MATRIX PRODUCTS INVOLVING (X, X^T, X^*, X^{*T}) OF FINITE ORDER. In this section we will be concerned with matrix functions of the form

$$Y = B_1 \cdot Z_1 \cdot B_2 \cdot Z_2 \cdots B_n \cdot Z_n \cdot B_{n+1}$$

$n > o$, n an integer and the B_i scalar matrices. Functions of Y, the argument, will be examined, such as $f(Y)=Y$, and certain transcendentals identified below. The algorithms to be developed are divided into three classes, viz: those in which the $Z_i \varepsilon \{X, X^T\}$, where the $Z_i \varepsilon (X^*, X^{*T})$ and those arguments where the Z_i can be any of these matrices. It will be seen that, for each of the functions (mappings) listed (in each class), the partial derivative of the trace with respect to X^T has a common form, and that the different solutions merely require the substitution of a specified matrix into this form. The generalized equations will be specified, but the formal proof will not be given. These equations can easily be proved by induction, but only the induction step is shown. It will be clear from the forms of the first two classes how the results can be combined to solve for a generalized mixed product. This will be indicated and the general solution presented.

Finally, at the end is a table of the partials of the trace for some of the more commonly expected products.

The following set definitions will be helpful. Let

S_A = {B : the B are finite rectangular scalar matrices.}

S_x = {X : the X are full rank rectangular variable matrices. $X \varepsilon S_x \Rightarrow X^T \varepsilon S_x$}

S_x^* = {X* : the X* are the pseudoinverses of the elements S_x.}

S_Y = {Y : the Y are square matrices of the form

$$Y = \left(\prod_{i=1}^{n} B_i X_i \right) B_{n+1}$$

where n>o is an integer, $B_i \varepsilon S_A$, $X_i \varepsilon S_x$}

S_Y^* = {Y : the Y are square matrices of the form

$$Y = \left(\prod_{i=1}^{n} B_i X_i^* \right) B_{n+1}$$

where n>o is an integer, $B_i \varepsilon S_A, X_i^* \varepsilon S_x^*$}

S_Z = {Z : the Z are square matrices of the form

$$Z = \left(\prod_{i=1}^{n} B_i Z_i \right) B_{n+1}$$

where n>o is an integer, $B_i \varepsilon S_A$, $Z_i \ \varepsilon \ S_Y \cup S_Y^*$}

Class I differentials will be developed first in considerable detail using the argument:

$$Y = A\ X\ B\ X^T C\ X\ D;\ A,\ B,\ C,\ D\varepsilon S_A;\ X,\ X^T\ S_x;\ Y\varepsilon S_Y \qquad (1)$$

followed by a more abbreviated development for Class II differentials using the argument:

$$Y = A\ X^*\ B\ X^*\ C;\ A,\ B,\ C\varepsilon S_A,\ X^*\varepsilon S_x,\ Y\varepsilon S_Y^* \qquad (2)$$

Specific functions of Class III arguments will not be developed.

The following auxilliary definitions will be helpful

$$\begin{aligned} D_1 &\triangleq AdZ\ (BZ^T CZD) \\ D_2 &\triangleq (AZB)dZ^T\ (CZD) \\ D_3 &\triangleq (AZBZ^T C)\ dZD \end{aligned} \qquad (3)$$

where $Z\varepsilon S_x$ or $Z\varepsilon S_x^*$. The restriction or extension of the subscripts is obvious.

Let $S\varepsilon S_A$. Then (since $tr\ (A)=tr(A^T)$)

$$\begin{aligned} tr\ (SD_1) &= tr\ (BZ^T CZDSAdZ) \\ tr\ (SD_2) &= tr\ ((AZB)^T\ S^T (CZD)^T\ dZ) \\ tr\ (SD_3) &= tr\ (DSAZBZ^T CdZ) \end{aligned} \qquad (4)$$

Let $F_M \triangleq \{\pm SY,\ e^{\pm Y},\ \pm\ \sin Y,\ \pm\ \cos Y,\ \pm\ \sinh Y,\ \pm\ \cosh Y\ :\ S\varepsilon S_A,$ $Y\varepsilon S_Y \cup S_Y^*\}$ and M be a mapping such that $M : Y \rightarrow M(Y)\varepsilon F_M$. Call this function in F_M, Q, and in all cases, let q be defined as

$$q \triangleq \mathrm{tr}\,(Q)$$

Class I

$$Y = AXBX^TCXD \tag{5}$$

Let

$$Q \in F_M = SY,\ S \in S_A \tag{6}$$

Then

$$\frac{\partial q}{\partial X^T}\ dX \triangleq dQ = S(D_1+D_2+D_3) \tag{7}$$

where the D_i are given by Equation (3). Using trace properties, and substituting Equation (4) into Equation (7)

$$\mathrm{tr}\,(DQ) = \mathrm{tr}\,(SD_1) + \mathrm{tr}\,(SD_2) + \mathrm{tr}\,(SD_3) \tag{8}$$

$$= \mathrm{tr}[(BX^TCXDSA+(AXB)^TS^T(CXD)^T+DSAXBX^TC)dX]$$

Therefore

$$\frac{\partial q}{\partial X^T} = BX^TCXDSA + (CXD\ SAXB)^T + DSAXB\ X^TC \tag{9}$$

and if S=I

$$\frac{\partial q}{\partial X^T} = \frac{\partial(\mathrm{tr}Y)}{\partial X^T} = BX^T\ CXDA + (CXDAXB)^T + DAXBX^TC \tag{10}$$

Let $Q \in F_M = e^Y$. Then Q can be expanded in a series [5, 6, 7, in particular, theorem 4, p 46, ref 7] as

$$Q = e^Y = I + Y + \frac{1}{2!}Y^2 + \frac{1}{3!}Y^3 + \cdots \tag{11}$$

Then

$$dQ = D_1 + D_2 + D_3 + \frac{1}{2!} (D_1Y+D_2Y+D_3Y+YD_1+YD_2+YD_3)$$
$$\frac{1}{3!} (D_1Y^2+D_2Y^2+D_3Y^2+YD_1Y+YD_2Y+YD_3Y+Y^2D_1+Y^2D_2+Y^2D_3$$
$$+ \cdots \qquad (12)$$

and

$$tr(dQ)=tr \, [(I+Y+\frac{1}{2!}Y^2+\cdots)D_1] + tr \, [(I+Y+\frac{1}{2!}Y^2+\cdots)D_2]$$
$$+ \, tr \, [(I+Y+\frac{1}{2!}Y^2+\cdots)D_3] \qquad (13)$$

since $tr(D_1Y+YD_1)=2tr(YD_1)$, etc. Therefore

$$tr(DQ)= tr \, [e^Y(D_1+D_2+D_3)] \qquad (14)$$

Since $e^Y \varepsilon S_A$, identify S with e^Y and substitute into Equation (9). This gives

$$\frac{\partial q}{\partial X^T} = BX^TCXDe^YA + (CXDe^YAXB)^T + De^YAXBX^TC \qquad (15)$$

We might note that, due to the permutative aspects of the trace,

$$\frac{\partial q}{\partial X^T}$$

can be written in a variety of ways. The one chosen here appears most "natural" because the argument, Y, remains unchanged. However, suppose $Y \triangleq AXB$ and $Q \triangleq e^Y$, then the following expressions are equivalent

$$\frac{\partial q}{\partial X^T} = Be^{AXB} \, A(=Be^YA) = e^{BAX}BA = BA \, e^{XBA}$$

A similar set of equalities exist for each function in F_M. However, the representation is unique if the argument remains fixed. In the remainder of this section, only the form which preserves $S\varepsilon F_M$ will be considered.

Now, had Q been $e^{iY}=\cos Y+i\sin Y$, the terms in the expansion of dQ could be collected into real and imaginary parts. It can be easily seen that these sums can be written as [3]

$$\mathrm{tr}(dQ) = -\,\mathrm{tr}\,[(\sin Y)(D_1+D_2+D_3)] + i\,\mathrm{tr}\,[(\cos Y)(D_1+D_2+D_3)] \quad (16)$$

Let $q_1 \triangleq \mathrm{tr}(\cos Y)$ and $q_2 \triangleq \mathrm{tr}(\sin Y)$. Then if S is chosen as $S=-\sin Y$, Equation (9) becomes

$$\frac{\partial q_1}{\partial X^T} = -\,B\,X^T\,CXD\,(\sin Y)\,A - (CXD(\sin Y)\,AXB)^T - D(\sin Y)\,AXBX^TC \quad (17)$$

If S is chosen as: $S=\cos Y$, Equation (9) becomes

$$\frac{\partial q_2}{\partial X^T} = B\,X^TCXD\,(\cos Y)\,A + (CXD(\cos Y)\,AXB)^T + D(\cos Y)\,AXBX^TC \quad (18)$$

Similarly, if the expansions of $\mathrm{tr}(e^Y)$ and $\mathrm{tr}(de^Y)$ are grouped according to even and odd powers, $\partial/\partial X^T\,\mathrm{tr}(\cosh Y)$ and $\partial/\partial X^T\,\mathrm{tr}(\sinh Y)$ can be expressed by substituting $S=\sinh Y$ and $S=\cosh Y$, respectively, in Equation (9).

There is a more enlightening way the above derivations can be performed. Let

$$X_1 \triangleq AX$$

$$X_2^T \triangleq (XB^T)^T \quad (19)$$

$$X_3 \triangleq CX$$

then

$$Y = X_1X_2^TX_3D$$

Also, for any i=1,2,3,Y can be expressed as

$$Y = L_i X_i M_i$$

(except for a transpose on X_2) and

$$L_1 = I$$

$$M_1 = X_2^T X_3 D$$

$$L_2 = X_1$$

$$M_2 = X_3 D \quad \text{(using } X_2^T\text{)} \tag{20}$$

$$L_3 = X_1 X_2^T$$

$$M_3 = D$$

Then Equation (4) can be written as

$$tr(SD_1) = tr\ (M_1 SL_1 dX_1) = (M_1 SL_1 AdX)$$

$$tr(SD_2) = tr\ (M_2 SL_2 dX_2^T) = tr\ ((M_2 SL_2 B)^T dX) \tag{21}$$

$$tr(SD_3) = tr\ (M_3 SL_3 dX_3) = tr(M_3 SL_3 CdX)$$

Such expressions can be generalized as follows. Let $\alpha \underline{\Delta}$"1" or "T" used as a superscript. That is, for any matrix, A,

$$A^\alpha = \begin{cases} A \text{ if } \alpha = 1 \\ A^T \text{ if } \alpha = T \end{cases} \tag{22}$$

Define

$$Y \triangleq \left(\prod_{i=1}^{n} X_i^{\alpha} \right) B_{n+1} \tag{23}$$

where

$$X_i \triangleq B_i X$$

and

$$X_i^T \triangleq (XB_i^T)^T = B_i X^T, \; B_i \varepsilon S_A, \; i = 1,2,\cdots,n+1 \tag{24}$$

Also, for any i, i=1,2,···,n, Y can be defined as

$$Y = L_i X_i^{\alpha} M_i \tag{25}$$

where

$$L_i = \prod_{j=1}^{i-1} X_j^{\alpha}, \; M_i = \left(\prod_{j=i+1}^{n} X_j^{\alpha} \right) B_{n+1} \tag{26}$$

where $L_1 \triangleq I$ and $M_n \triangleq B_{n+1}$. Then for $S \varepsilon S_A$ and $Q=M(Y)\varepsilon F_m$

$$tr(dQ) = tr \left[\sum_{i=1}^{n} (M_i S L_i B_i)^{\alpha} dX \right] \tag{27}$$

which gives

$$\frac{\partial q}{\partial X^T} = \sum_{i=1}^{n} (M_i S L_i B_i)^{\alpha}, \; S \varepsilon F_M \tag{28}$$

Note that if $M(Y)\varepsilon\{ TY, e^{Y} : T\varepsilon S_A \}$ then S=M(Y); if M(Y)=cosY, then S=-sinY; if M(Y)=sinY, then S=cosY; if M(Y)=sinhY, then S=coshY; and if M(Y)=coshY, S=sinhY.

Class II. The development of a generalized form using the pseudo-inverse parallels the previous derivations very closely; only the form of the final result is changed. Therefore, it will be sufficient for illustrative purposes to use Y as defined in Equation (2)

$$Y = AX^*BX^*C;\ A,B,C\varepsilon S_A,\ X^*\varepsilon S^*_X,\ Y\varepsilon S^*_Y$$

By assumption X is full rank, so Equations (114) through (117) of the previous section apply. Some of these are

$$X^* = (X^TX)^{-1}X^T$$

$$G \triangleq (X^TX) \Rightarrow X^* = G^{-1}X^T$$

$$X^*X = I \tag{29}$$

$$XX^* \triangleq P, \text{ a projector}$$

and $G=G^T$, $P=P^T$. Let

$$D_1 \triangleq AdX^*BX^*C \tag{30}$$

$$D_2 \triangleq AX^*BdX^*C$$

Then for $S\varepsilon S_A$

$$tr(SD_1) = tr(BX^*CSAdX^*)$$

$$tr(SD_2) = tr(CSAX^*BdX^*) \tag{31}$$

We require dX^*. From Equation (29)

$$dX^* = dG^{-1}X^T + G^{-1}dX^T,$$

From equation (62) of the previous section

$$dG^{-1} = -\ G^{-1}dGG^{-1}$$

and from

$$G = X^T X$$

$$dG = dX^T X + X^T dX$$

Thus

$$dX* = - G^{-1} dX^T X G^{-1} X^T - G^{-1} X^T dX G^{-1} X^T + G^{-1} dX^T \tag{32}$$

If $T \epsilon S_A$, then

$$\begin{aligned} tr(TdX*) &= - tr(TG^{-1} dX^T X G^{-1} X^T) - tr(TG^{-1} X^T dX G^{-1} X^T) \\ &\quad + tr(TG^{-1} dX^T) \\ &= - tr(TG^{-1} dX^T XX*) - tr(TX* dXX*) + tr\ (TG^{-1} dX^T) \\ &= - tr(PdXG^{-1} T^T) - tr(X*TX*dX) + tr\ (dXG^{-1} T^T) \\ &= tr(G^{-1} T^T (I-P)dX) - tr(X*TX*dX) \end{aligned} \tag{33}$$

Thus, if $Q=SY$, $S \epsilon S_A$, define

$$\begin{aligned} A_1 &\triangleq BX*CSA \\ A_2 &\triangleq CSAX*B \end{aligned} \tag{34}$$

Then

$$\begin{aligned} tr(dQ) &= tr(SD_1) + tr(SD_2) = tr(A_1 dX*) + tr(A_2 dX*) \\ &= tr[G^{-1}(A_1^T + A_2^T)(I-P)dX] - tr[X*(A_1 + A_2)X*dX] \end{aligned} \tag{35}$$

or

$$\frac{\partial q}{\partial X^T} = G^{-1}(A_1^T + A_2^T)(I-P) - X*(A_1 + A_2)X*$$

We note that if $Q \epsilon F_M$, then $S \epsilon F_M$ according to the scheme outlined for Equation (28).

The above procedure can be generalized in the same manner as before, where "α" as a superscript has the same meaning as in Equation (22). Let

$$Y \triangleq \left(\prod_{i=1}^{n} (X_i^*)^\alpha \right) B_{n+1} \tag{36}$$

$$X_i^* = B_i X^*$$

$$(X_i^*)^T \triangleq B_i (X^*)^T = (X^* B_i^T)^T, \; B_i \varepsilon S_A, \text{ and } X^* \varepsilon S_x^* \tag{37}$$

Then, for any i, Y can be defined as

$$Y \triangleq L_i (X_i^*)^\alpha M_i. \tag{38}$$

where L_i and M_i have the same definitions as in Equation (26) with X_i. It follows that

$$tr(dQ) = tr \left[\sum_{i=1}^{n} (M_i S L_i B_i)^\alpha dX^* \right] \tag{39}$$

Now define

$$\Gamma \triangleq \sum_{i=1}^{n} (M_i S L_i B_i)^\alpha \tag{40}$$

then

$$tr(dQ) = tr(\Gamma dX^*) = tr[(G^{-1}\Gamma^T(I-P)-X^*\Gamma X^*)dX] \tag{41}$$

Thus

$$\frac{\partial q}{\partial X^T} = G^{-1}\Gamma^T(I-P) - X^*\Gamma X^* \tag{42}$$

and the same remarks apply as for Equation (28).

Class III. From the foregoing derivations it is clear that if

$$Y = \left(\prod_{i=1}^{n} Z_i \right) B_{n+1}, \quad Y \varepsilon S_Z$$

where Z_i can be either $B_i X^\alpha$ or $B_i (X^*)^\alpha$, then for any i

$$Y = L_i Z_i M_i$$

and, if Z' is X or X^*, then

$$tr(dQ) = tr \left[\sum_{i=1}^{n} (M_i S L_i B_i)^\alpha dZ' \right]$$

$$= tr \left[\sum_{k=1}^{n_1} (M_k S L_k B_k)^\alpha dX \right] + tr \left[\sum_{m=1}^{n_2} (M_m S L_m B_m)^\alpha dX^* \right] \quad (43)$$

where k sums over the n_1 terms involving dX and m sums over the n_2 terms involving dX^*; the M_i, L_i have the same definitions as before. Therefore, if we define

$$\Gamma_1 = \sum_{k=1}^{n_1} (M_k S L_k B_k)^\alpha$$

$$\Gamma_2 = \sum_{m=1}^{n_2} (M_m S L_m B_m)^\alpha \quad (44)$$

$$\boxed{\frac{\partial q}{\partial X^T} = \Gamma_1 + G^{-1} \Gamma_2^T (I-P) - X^* \Gamma_2 X^*} \quad (45)$$

Again $S \varepsilon F_M$ and the remarks given for Equation (28) apply.

The general forms given above used the symbol "S" to indicate a matrix function in the resultant expression. Furthermore, the form of the result was restricted among several possibilities so that $S \varepsilon F_M$. This restriction assures the uniqueness of the result. More formally, S can be defined as follows. Let $Q \varepsilon F_M$, q=trQ, then

$$S = \frac{\partial q}{\partial Y} \quad (46)$$

For example, if $Q=e^{Y}$, then

$$dQ = \frac{\partial q}{\partial Y} dY$$

and

$$\text{tr } dQ = \text{tr}[d(e^{Y})] = \text{tr}[d(I+Y+\frac{Y^2}{2!}+\frac{Y^3}{3!}+\cdots)]$$

$$= \text{tr}[(I+Y+\frac{Y^2}{2!}+\cdots)dY] = \text{tr}[e^{Y}dY] \qquad (47)$$

and $\partial q/\partial Y=e^{Y}$. Clearly, comparable results follow for each element in F_M.

Throughout the development which resulted in the general equation (45), attention has been confined to the functions in F_M. It is clear that any function expressible as a convergent series has a matrix counterpart [7, theorem 4, p 46]. Thus, the set F_M could be expanded considerably; it would include, among others, such functions as logY, Bessel functions, etc., in which S can be substituted as shown in Equations (44), based on the result demonstrated by (47), $S \epsilon F_M$.

Based on (47), the derivative of f(Y) with respect to Y is the same as if Y were a scalar. Thus, if Q=sinY,

$$\frac{\partial q}{\partial Y} = \frac{\partial}{\partial Y} \text{tr}(\sin Y) = \cos Y, \text{ etc.}$$

Of particular importance in applications are the matrix functions AXB and AX^*B. Table 1, below lists the gradient of the trace, $\partial q/\partial X^T$ of these two forms for the functions defined in F_M. Capital letters represent matrices in S_A, with $X \epsilon S_X$ and $X^* \epsilon S_X^*$.

TABLE 1

Y	Q	$\partial q/\partial X^T$
Y=ABX	Y	BA
	CY	BCA
	e^Y	Be^YA
	sinY	B(cosY)A
	cosY	-B(sinY)A
	sinhY	B(coshY)A
	coshY	B(sinhY)A
Y=AX*B[1]	Y	$G^{-1}A^TB^T(I-P)-X^*BAX^*$
	CY	$G^{-1}A^TC^TB^T(I-P)-X^*BCAX^*$
	e^Y	$G^{-1}A^T(e^Y)^TB^T(I-P)-X^*Be^YAX^*$
	sinY	$G^{-1}A^T(\cos Y^T)B\ (I-P)-X^*B(\cos Y)AX^*$
	cosY	$-G^{-1}A^T(\sin Y^T)B^T(I-P)+X^*B(\sin Y)AX^*$
	sinhY	$G^{-1}A^T(\cosh Y^T)B^T(I-P)-X^*B(\cosh Y)AX^*$
	coshY	$G^{-1}A^T(\sinh Y^T)B^T(I-P)-X^*B(\sinh Y)AX^*$

[1] $G=(X^TX)$, $P=XX^*$

More generality is available using Equations (26), (44), and (45). Table 2 lists the functions $Q\epsilon F_M$, and $S=\partial q/\partial Y$ which are required in the gradient of the trace of Q, $\partial q/\partial X^T$, for each choice of Q. Table 3 lists a number of forms for Y and $\partial q/\partial X^T$ in which the required function, S, is found in Table 2 depending on Q.

TABLE 2

Q	AY	e^Y	cosY	sinY	coshY	sinhY
S	A	e^Y	-sinY	cosY	sinhY	coshY

TABLE 3

Y	$\partial q/\partial X^T$
AXB	BSA
AXBXC	BXCSA + CSAXB
$AXBX^TC$	$BX^TCSA + (CSAXB)^T$
XX^T	$X^T(S+S^T)$
XBX^T	$BX^TS + B^TX^TS^T$
AXX^TB	$X^TBSA + (BSAX)^T$
AXXB	XBSA + BSAX
AX*B	$G^{-1}(BSA)^T(I-P) - X^*BSAX^*$
AX*BX*C	$G^{-1}\Gamma^T(I-P)-X^*\Gamma X^*$, where $\Gamma = BX^*CSA + CSAX^*B$
$AX^*BX^{*T}C$	$= G^{-1}\Gamma^T(I-P) - X^*\Gamma X^*$, where $\Gamma = BX^{*T}CSA + (CSAX^*B)^T$
X^*X^{*T}	$= G^{-1}(S+S^T)X^*(I-P) - X^*X^{*T}(S+S^T)X^*$
X^*X^*	$= G^{-1}(S^TX^{*T}+X^{*T}S^T) - X^*(X^*S+SX^*)X^*$
$X^{*T}X^{*T}$	$= G^{-1}(X^{*T}S+SX^{*T}) - X^*(S^TX^*+X^*S^T)X^*$
AX*BXC	$= CSAX^*B + G^{-1}\Gamma^T(I-P) - X^*\Gamma X^*$, where $\Gamma = BXCSA$
AX^*BX^TC	$= (CSAX^*B)^T + G^{-1}\Gamma^T(I-P) - X^*\Gamma X^*$, where $\Gamma = BX^TCSA$
$AX^{*T}BXC$	$= CSAX^{*T}B + G^{-1}\Gamma^T(I-P) - X^*\Gamma X^*$, where $\Gamma = (BXCSA)^T$

IV. APPLICATIONS TO KALMAN FILTERING THEORY. This section presents an application of gradient techniques to the continuous Kalman filter.

The optimal state vector defined by the Kalman filter, depending on context, can be considered as an estimation of the state or as a filtered variable. Although there are many ways one can formulate estimation or filtering equations, the Kalman formulation is sufficiently general so that a considerable variety of such problems are expressible in that manner. The derivation of the optimal gain matrix is ideally suited to demonstrate the use of the techniques developed in this paper.

The presentation, below, omits most details of the theory already extensively recorded in the open literature; sufficient detail is presented so as to obtain the matrix Riccati equation for the propagation of the covariance of the state.

Consider a "truth model"

$$\dot{x}(t)(n\rangle_j = A\, x(t)(n\rangle_j + f(t)(n\rangle + \underset{n\times\ell}{B}\, u(t)(\ell\rangle \tag{1}$$

for a countably infinite sequence of continuous trajectories, $j=1,\cdots,\infty$. The measurement vector is

$$z(t)(m\rangle_j = \underset{m\times n}{A}(t)\, x(t)(m\rangle_j + v(t)(m\rangle_j \tag{2}$$

The structure of the optimal state estimate is

$$\dot{\hat{x}}(t,t)\rangle_j = A\hat{x}(t,t)\rangle_j + f(t)\rangle + W(t)\,\tilde{x}(t,t)\rangle_j \tag{3}$$

where the error in observation is

$$\tilde{z}(t,t)_j = z(t)_j - \hat{z}(t,t)_j \tag{4}$$

The error dynamics are

$$\dot{\tilde{x}}(t,t)\rangle_j = (A-WH)\,\tilde{x}(t,t)\rangle_j + Bu\rangle_j + Wv\rangle_j \tag{5}$$

The variance in the state estimation is

$$P(t,t) = E(\tilde{x}(t,t)\rangle\langle\tilde{x}(t,t)) \tag{6}$$

with dynamics given by

$$\dot{P} = (A-WH)P + P(A^T - H^T W^T) + BQB^T + WRW^T \tag{7}$$

where the standard uncorrelated assumptions are assumed and

$$E\,(u\rangle\langle u) = Q(t,\tau)\,\delta(t,\tau)$$

and

$$E\,(v\rangle\langle v) = R(t,\tau)\,\delta(t,\tau)$$

Any full rank W used in Equation (3) will generate a trajectory; the problem is to find the W(t) that will minimize the trace of Equation (7). It can be shown that if $\dot{P}$ is minimized, then P is minimized, hence

$$\frac{\partial \dot{P}}{\partial W^T} = \frac{\partial}{\partial W^T}(-WHP) - \frac{\partial}{\partial W^T}(PH^T W^T) + \frac{\partial}{\partial W^T}(WRW^T) \tag{8}$$

From Table 3, Section III

$$\frac{\partial}{\partial W^T}\,\mathrm{tr}\,(-WHP) = -\,HP \tag{9}$$

$$\frac{\partial}{\partial W^T}\,\mathrm{tr}\,(PH^T W^T) = +\,HP \tag{10}$$

$$\frac{\partial}{\partial W^T}\,\mathrm{tr}\,(WRW^T) = 2RW^T \tag{11}$$

Using Equations (9) through (11) in Equation (8), gives

$$\frac{\partial \dot{P}}{\partial W^T} = 0 = -\,2HP + 2RW^T$$

from which

$$W^T = R^{-1}HP$$

That is,

$$\boxed{W = PH^T R^{-1}} \qquad (12)$$

Which is the standard Kalman gain matrix.

V. REFERENCES

1. Athans, M., Schweppe, F.C. "Gradient Matrices and Matrix Calculations". MIT Lincoln Laboratory, Technical Note, 1965-53, 17 Nov 1965.

2. Tracy, D.S. and Dwyer, P.S., "Multivariate Maxima and Minima with Matrix Derivatives". American Statistical Journal, Dec 1969. pp 1576.

3. Neudecker, H., "On Matrix Procedures for Optimizing Differentiable Scalar Functions of Matricies". Statistica Neerlandica, 21 (1967), Nr 1.

4. Pappas, J.S., "A Tutorial Derivation of Recursive Weighted Least Squares State - Vector Estimation Theory (Kalman Theory)". US Army Test and Evaluation Command, Analysis and Computation Directorate, White Sands Missile Range, NM. Aug 1968.

5. Pipes, Louis A., "Matrix Methods for Engineering", Prentice-Hall, Inc., Englewood Cliffs, NJ, 1963. Chap 4.

6. Meditch, J.S., "Stochastic Optimal Linear Estimation and Control", McGraw-Hill Book Co, NY, NY, 1969. pp 24-51.

7. Pullman, J. J., "Matrix Theory and its Applications", Marcel Dekker, Inc, NY and Basel, 1976. pp 35-62.

APPENDIX

NOTATION

1. Upper case Roman letters A, B, C, ···, generally refer to rectangular matrices.

Special Forms:

- I — Is a unit matrix; I_p a p×p unit matrix.
- O — Is a matrix of zeros; O_p a p×p matrix of zeros.
- X — A rectangular matrix of random variables or a set of measurements. Each column is a measurement and the number of rows is generally greater than or equal to the number of columns. X is normally "full-rank", that is, the rank of X equals the number of columns.
- X^* — Is the generalized inverse of X. In particular, it is the pseudo inverse of X when X is assumed to be full-rank. That is, if the rank of X is equal to the number of its columns, then $X^*=(X^TX)^{-1}X^T$. (NOTE: If the rank of X were equal to the numbers of its rows, then $X^*=X^T(XX^T)^{-1}$.) If X is square, then $X^*=X^{-1}$.
- G — Is the Grammian of a full row rank matrix; thus $G=X^TX$.
- W — Generally refers to a matrix of weights.
- P,Q,R — Frequently refer to the covariances of the optimal estimate of the state-vector, process noise, and measurement noise, resp.
- $P,\tilde{P}$ — Represent a projector of a variable array into its range space and null space, respectively. Both P and $\tilde{P}$ are idempotent and $P\tilde{P}=0$.

2. Lower case Roman letters, a, b, c, ···, refer to vectors.

Special Forms. Frequently an upper case Roman letter, say A, will refer to an unspecified array. For those special cases in which the array is to be a vector, lower case a will replace A.

3. Operators.

Superscripts

T — Is the transpose of an array; thus A^T.

-1 — Is the inverse of a square array; thus A^{-1}.

* — Is the pseudo (generalized) inverse of an array; thus A^*.

(If A has full-rank, $A^* = (A^T A)^{-1} A^T$.)

Symbols

tr — Is the trace of a square array; thus $trA = \sum_{i=1}^{n} a_{ii}$ where $A = (a_{ij})$ is an n×n matrix.

$\langle , \rangle$ — The Dirac bra and ket, respectively, represent row and column vectors respectively. Since, in this paper, all arrays are of real numbers, only, the inner product of two vectors a and b is $\langle ab\rangle = \langle ba\rangle$ or $\langle aMb\rangle = \langle bM^T a\rangle$ over the metric array, M.

$\langle\!\langle pa$, $bq\rangle\!\rangle$ — Represent row and column vectors a and b, having p and q compontents, respectively.

$a\rangle\!\rangle_\alpha$, ${}^\beta\langle\!\langle b$ — α,β natural numbers, are used to represent arrays A and B partitioned into their column space or row space, respectively, or are used to emphasize that two vectors, a and b, are defined in a "space" and its "cospace" defined by a "basis" and "reciprocal basis".

^, ~ — Appearing above an array respresent an optimal estimate and its optimal error, respectively, thus $\hat{X}$ and $\tilde{X}$. Also note that such an array can be written as $X = \hat{X} + \tilde{X}$.

E(·), cov(·) — Represent expected value and covariance of the arguments, respectively.

DISTRIBUTION LIST:

Defense Documentation Center Cameron Station Alexandria, Virginia 22314	12 cys

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Analysis and Computation Division National Range Operations Directorate White Sands Missile Range, NM 88002	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP NA

3. REPORT TITLE

NEW GRADIENT TECHNIQUES FOR TRACES OF FUNCTIONS OF RECTANGULAR MATRICES AND THEIR PSEUDO-INVERSES

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (First name, middle initial, last name)

James S. Pappas and Oren N. Dalton

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
September 1977		6

8a. CONTRACT OR GRANT NO. NA	9a. ORIGINATOR'S REPORT NUMBER(S) NA
b. PROJECT NO. NA	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	ACD-61-77

10. DISTRIBUTION STATEMENT

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY

13. ABSTRACT

The generalized inverse is of increasing importance for estimation and optimization in modern systems theory, because it both simplifies many problems and reveals underlying structures of theoretical importance. Optimization, using the gradient of the trace of products of matrix valued functions, including the generalized inverse, are presented in a novel state-space setting. A number of functions of such products of matrices, including general formulas, are derived for the first time to our knowledge. Tables of a number of them are given with some applications in the fields of estimation and optimization.

DD FORM 1 NOV 65 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Gradient Matrices Optimization Generalized Inverse Filtering Theory						