# Matrix Algebra, Class Notes (part 2)

by

# Hrishikesh D. Vinod



## 1 Converting Matrices Into (Long) Vectors

**Convention: Let A be a T**$\times m$ **matrix; the notation vec(A) will mean the Tm-element column vector whose first set of T elements are the first colum of A, that is a**$_{\cdot 1}$ **using the dot notation for columns; the second set of T elements are those of the second column of A, a**$_{\cdot 2}$**, and so on. Thus** $\mathbf{A} = [\mathbf{a}_{\cdot 1},\ \mathbf{a}_{\cdot 2}, \cdots, a_{\cdot m}]$ **in the dot notation.**

**An immediate consequences of the above Convention is Vec of a product of two matrices contains a Kronecker with the identity, (remember to transpose and write the second matrix before the kronecker).**

**Exercise: Let A, B be T**$\times m$**, m**$\times q$ **respectively. Then using the Kronecker product notation, we have**

$$\mathbf{vec(AB)} = (\mathbf{B}\ ' \otimes I)\ \mathbf{vec(A)} = (\mathbf{I} \otimes \mathbf{A})\ \mathbf{vec(B)} \tag{1}$$

**For example, if A=** $\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}$ **and B=** $\begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix}$ **then AB=** $\begin{bmatrix} a_1b_1 + a_2b_3 & a_1b_2 + a_2b_4 \\ a_3b_1 + a_4b_3 & a_3b_2 + a_4b_4 \end{bmatrix}$

**vec(AB)=** $[] = (B\ ' \otimes I)$ **vec(A)**

$= \begin{bmatrix} b_1\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & b_3\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ b_2\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & b_4\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{bmatrix}$ [] **times** $Vec(A)$ **can then be verified.**

### 1.1 Proposition 1

**Vec of a Product of Three Matrices has Kronecker with Identity unless the vec of the middle matrix is used: Let** $A_1,\ A_2,\ A_3$ **be conformably dimensioned matrices. Then**

$$\mathbf{vec}(A_1A_2A_3) = (\mathbf{I} \otimes A_1A_2)\ \mathbf{vec}(A_3) \tag{2}$$

$= (\mathbf{A}_3' \otimes A_1)\ \mathbf{vec}(A_2)$

$= (\mathbf{A}_3' A_2' \otimes \mathbf{I})$ **vec**$(A_1)$.

## 1.2 Proposition 2

**Vec of a Sum of Two Matrices is the sum of vecs. Let A, B be T**$\times n$**. Then**
**vec(A + B) = vec(A) + vec(B).** **(3)**

**Corollary: Let A, B, C, D be conformably dimensioned matrices. Then from (1)**
**vec[(A + B)(C + D)] = [(I**$\otimes$ **A) + (I** $\otimes$ **B)][vec(C) + vec(D)]** **(4)**

$= [(C\,'$ $\otimes$ **I) + (D** $'$ $\otimes$ **I)][vec(A) + vec(B)]**

## 1.3 Proposition 3

**Trace of a Product of Two Matrices in the Vec notation (prime of the vec of a prime). Let A, B be conformably dimensioned matrices. Then**
**tr(AB) = vec(A** $'$ **)** $'$ **vec(B) = vec(B** $'$ **)**$'$ **vec(A).** **(5)**

**For above example, tr(AB)=** $(a_1b_1 + a_2b_3) + (\ a_3b_2 + a_4b_4) = [\ a_1\ \ a_2\ \ a_3\ \ a_4]\ []$

**illustrates the first part of the proposition 3 regarding the trace.**

## 1.4 Proposition 4

**Trace of a product of three matrices involves the prime of a vec and a kronecker with the identity. Let** $A_1$, $A_2$, $A_3$ **be conformably dimensioned matrices. Then**
**tr**$(A_1A_2A_3)$ = **vec**$(A_1)\ '\ (A_3 \otimes$ **I)vec**$(A_2)$ **(6)**

$=$ **vec**$(A_1)\ '$ **(I** $\otimes A_2) vec(A_3)$
$=$ **vec**$(A_2)\ '$ **(I** $\otimes A_3)$ **vec**$(A_1)$
$= vec(A_2)\ '\ (A_1 \otimes I)$ **vec**$(A_3)$
$=$ **vec**$(A_3)\ '\ (A_2 \otimes I)$ **vec**$(A_1)$

## 2 Basics of Vector and Matrix Differentiation

**In the derivation of the least squares we minimize the error sum of squares by using the first order conditions, i.e., we differentialte the error sum of squares. This can be formulated in matrix notation. Sometimes we need to differentiate quantities like tr(AX) with respect to the elements of X, or quantities like Ax, z**$'Ax$ **with respect to the elements of (the vectors) x and/or z. Although no fundamentally new concept is involved in carrying out such matrix differentiations, they can seem to be cumbersome. If we have simple rules and conventions, matrix differentiation becomes a powerful tool for research.**
**Convention of using the same numerator subscript along each row. Let y =** $\psi(x)$, **where y is a T**$\times 1$ **vector and x is an n**$\times 1$ **vector. The symbol**

$$\frac{\partial y}{\partial x} = \frac{\partial y_i}{\partial x_j} \text{ , i = 1,2,}\cdots, T, \text{ and j = 1,2}\cdots, n \tag{1}$$

**denotes the T**$\times n$ **matrix of first-order partial derivatives, the so-called Jacobian matrix of the transformation from x to y. For example, if y**$' = (y_1 \ y_2 \ y_3)$ **and x**$' = (x_1 \ x_2)$ **the convention states that:**

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \partial y_1/\partial x_1 & \partial y_1/\partial x_2 \\ \partial y_2/\partial x_1 & \partial y_2/\partial x_2 \\ \partial y_3/\partial x_1 & \partial y_3/\partial x_2 \end{bmatrix} \text{ is, in general, a T}\times n \text{ matrix.} \tag{1}$$

**Observe that ith row contains the derivatives of the ith element of y with respect to the elements of x. The numerator is y and its subscript is fixed along each row according to the convention. When both x and y are vectors, the** $\partial y/\partial x$ **has as many rows are are the rows in y and as many columns as are the rows in x. In particular, if y is 1**$\times 1$ **(i.e., a scalar), then the above Convention implies that** $\partial y/\partial x$ **is a row vector. If we wish to represent it as a column vector we may do so by writing** $\partial y/\ \partial x'$ **, or** $(\partial y/\ \partial x)'$ **. On the other hand, if x is a scalar** $(\partial y/\ \partial x)$ **is a column vector.**

**In general, if Y is a T**$\times p$ **matrix and x is an n**$\times 1$ **vector,** $\partial Y/\partial x$ **is a Tp**$\times n$ **matrix**
$$\frac{\partial Y}{\partial x} = \frac{\partial}{\partial x} \text{ vec(Y).} \tag{2}$$

where $\partial/\partial x$ makes an n$\times 1$ ROW vector of the same dimension as x. In the context of Taylor series and elsewhere it may be convenient to depart from this convention and let $\partial/\partial x$ be a column vector.

We now state some useful results involving matrix differentiation. Instead of formal proofs, we indicate some analogies with the traditional calculus and make some other comments to help the reader in remembering the results.

In (scalar) calculus if y$= 3x$, then $\partial y/\partial x = \mathbf{3}$. Similarly, in matrix algebra Derivative of a matrix times a vector is the matrix: Let y be a T$\times 1$ vector, x be an n$\times 1$ vector, and let **A** be a T$\times n$ matrix which does not depend on x. We have

y $=$ **Ax**, implies $(\partial y/\partial x) = \mathbf{A}$. **(3)**

Note that the i-th element of y is given by $y_i = \sum_{k=1}^{n} a_{ik}\ x_k$. Hence, $\frac{\partial y_i}{\partial x_j} = a_{ij}$ .

Exercise 1: If y$= a'x = x'a$, where a and x are T$\times 1$ vectors show that $\partial y/\partial x$ is a row vector according to our convention, and it equals a$'$.

Exercise 2: If S$= y'y$, where y is a T$\times 1$ vector show that $\partial S/\partial \beta = 0'$, where $\beta$ and **0** are p$\times 1$ vectors, since S is not a function of $\beta$ at all. This is useful in the context of minimizing the error sum of squaes u$'u$ in the regression model y$= X\beta + u$.

Exercise 3: If S$= y'X\beta$, where y is a T$\times 1$ vector, **X** is a T$\times p$ matrix, and $\beta$ is a p$\times 1$ vector show that $\partial S/\partial \beta = y'X$, This is also used in minimizing the residual sum of squaes in regression.

Chain Rule in Matrix Differentiation: In calculus if y$= f(x)$, and x$= g(\theta)$ then $\partial y/\partial \theta = (\partial y/\partial x)(\partial x/\partial \theta)$. Similarly in matrix algebra we have the following.

If y $=$ **Ax**, as in (3) above, except that the vector x is now a function of another set of variables, say those contained in the r-element column vector $\theta$, then we have

(the T$\times r$ matrix) $\frac{\partial y}{\partial \theta} = \frac{\partial y}{\partial x}\ \frac{\partial x}{\partial \theta} = \mathbf{A}\ \frac{\partial x}{\partial \theta}$ (where $\frac{\partial x}{\partial \theta}$ is an n$\times r$ matrix) **(4)**

Convention for second order derivatives (vec the matrix of first partials before taking second partials). Let y $= \psi(x)$ where y is a T$\times 1$ vector and x is an n$\times 1$ vector. Now by the symbol $\partial^2 y/\ \partial x\ \partial x'$ we shall mean

**(the Tn$\times n$ matrix)** $\frac{\partial^2 y}{\partial x \partial x'} = \frac{\partial}{\partial x}$ **vec**[ $\frac{\partial y}{\partial x}$ ] **(5)**

**so that the second order partial is a matrix of dimension (Tn)$\times n.$ Operationally one has to first convert the T$\times n$ matrix of first partials illustrated by (1) above into a long Tn$\times 1$ vector, and then compute the second derivatives of each element of the long vector with respect to the n elements of x written along the n columns$-$*giving* n columns for each of the Tn rows.**
**Chain Rule forSecond order partials w.r.t. $\theta$. Let y $=$ Ax be as in (3) above. Then**

$\frac{\partial^2 y}{\partial \theta \partial \theta'} =$ **(A** $\otimes$ $I_r$) $\frac{\partial^2 x}{\partial \theta \partial \theta'}.$ **(6)**

**Exercise: True or False ?** $\partial^2 y/( \ \partial\theta \ \partial\theta \ ' \ )$ **is of dimension Tnr$\times nr,$ (A $\otimes$ $I_r$) is of dimension Tn$\times n$ and** $\partial^2 x/ \ (\partial\theta \ \partial\theta \ ' \ )$ **is nr$\times r$ .**
**First order partial, x and A are functions of $\theta :$ Let y $=$ Ax, where y is T$\times 1,$ A is T$\times n,$ x is n$\times 1,$ and both A and x depend on the r-element vector $\theta.$ Then the T$\times r$ matrix**
$\frac{\partial y}{\partial \theta} =$ **(x** $'$ $\otimes$ $I_T$) $\frac{\partial A}{\partial \theta}$ + **A** $\frac{\partial x}{\partial \theta}.$ **(7)**

**where $\partial A/\partial\theta$ is Tn$\times r.$ Next we consider the differentiation of bilinear and quadratic forms.**

**Derivative of a Bilinear form involves first vector prime times the matrix (or the second vector prime times the matrix prime.) Let y $=$ z $'Ax,$ where y is a scalar, z is T$\times 1,$ A is T$\times n,$ x is n$\times 1,$ and A is independent of z and x. Now the 1$\times T$ vector of derivatives is as follows.**
$\frac{\partial y}{\partial z} =$ **x** $'A',$ **and the 1$\times n$ vector** $\frac{\partial y}{\partial x} =$ **z** $'A.$

**First Derivative of a Quadratic form. Let**
**y** $=$ **x** $'Ax,$
**where x is n$\times 1,$ and A is n$\times n$ square matrix and independent of x. Then a 1$\times n$**

**vector**
$\frac{\partial y}{\partial x}$ = **x** $'(A +$ **A** $')$.

**Exercise 1: If A is a symmetric matrix, show that** $\partial y/\partial x = 2$ **x**$'A$.
**Exercise 2: Show that** $\partial(\ \beta' X' X\beta)/\partial\beta = 2\beta' X' X$.
**Exercise 3: In a regression model y**$=X\beta + u$, **minimize u**$'u$ **and show that a necessary condition for a minimum is 0**$=$ **2**$\beta' X' X - 2y' X$, **and solve for** $\beta$.
**Second Derivative of a Quadratic form. Let A, y, and x be as in Proposition 6 above, then**

$$\frac{\partial^2 y}{\partial x \partial x'} = \mathbf{A}' + \mathbf{A},$$

**and, for the special case where A is symmetric,** $\frac{\partial^2 y}{\partial x \partial x'} = \mathbf{2A}$.

**Derivatives of a Bilinear Form with respect to** $\theta$. **If y** $=$ **z** $'Ax$,
**where z is T**$\times 1$, **A is T**$\times n$, **x is n**$\times 1$, **and both z and x are a function of the r-element vector** $\theta$, **while A is independent of** $\theta$. **Then** $\frac{\partial y}{\partial \theta}$ = **x** $'A\ '\ \frac{\partial z}{\partial \theta}$ + **z** $'A\ \frac{\partial x}{\partial \theta}$

**where** $\frac{\partial y}{\partial \theta}$ **is 1**$\times r$, $\frac{\partial z}{\partial \theta}$ **is T**$\times r$ **and** $\frac{\partial x}{\partial \theta}$ **is n**$\times r$. **Now the second derivcative is as follows.**

$$\frac{\partial^2 y}{\partial \theta \partial \theta'} = [\ \frac{\partial z}{\partial \theta}]\ '\ \mathbf{A}\ \frac{\partial x}{\partial \theta} + [\ \frac{\partial x}{\partial \theta}]\ '\ \mathbf{A}\ '\ [\frac{\partial z}{\partial \theta}] + (\mathbf{x}\ 'A\ ' \otimes \mathbf{I})\ \frac{\partial^2 z}{\partial \theta \partial \theta'}$$

$$+ (\mathbf{z}\ 'A \otimes \mathbf{I})\ \frac{\partial^2 x}{\partial \theta \partial \theta'}.$$

**Derivatives of a Quadratic Form with respect to** $\theta$.. **Consider the quadratic form y** $=$ **x**$'Ax$
**where x is n**$\times 1$, **A is n**$\times n$, **and x is a function of the r-element vector** $\theta$, **while A is independent of** $\theta$. **Then** $\frac{\partial y}{\partial \theta} = \mathbf{x}'(A' + \mathbf{A})\ \frac{\partial x}{\partial \theta}$,

$$\frac{\partial^2 y}{\partial \theta \partial \theta'} = (\ \frac{\partial x}{\partial \theta}\ )'\ (\mathbf{A}' + \mathbf{A}\ (\frac{\partial x}{\partial \theta}) + (\mathbf{x}'[A' + \mathbf{A}] \otimes \mathbf{I})\ \frac{\partial^2 x}{\partial \theta \partial \theta'}.$$

**Derivatives of a Symmetric Quadratic Form with respect to** $\theta$. **Consider the same situation as in Corollay above but suppose in addition that A is symmet-**

**ric. Then**
$\frac{\partial y}{\partial \theta} = \mathbf{2x}'A\ \frac{\partial x}{\partial \theta}$,

$$\frac{\partial^2 y}{\partial \theta \partial \theta'} = \mathbf{2}\ (\ \frac{\partial x}{\partial \theta}\ )'\ \mathbf{A}\ (\ \frac{\partial x}{\partial \theta}\ ) + (\mathbf{2x}'A \otimes \mathbf{I})\ \frac{\partial^2 x}{\partial \theta \partial \theta'}.$$

**First Derivative of a Bilinear form w.r.t the matrix. Let**
$\mathbf{y} = \mathbf{a}\ 'Xb$,
**where a and b are n$\times$1 vectors of constants and X is an n$\times n$ square matrix. Then an n$\times n$ matrix $\frac{\partial y}{\partial X} = \mathbf{ab}\ '$.**

**First Derivative of a Quadratic form w.r.t the matrix. Let**
$\mathbf{y} = \mathbf{a}\ 'Xa$,
**where a is an n$\times$1 vector of constants and X is an n$\times n$ symmetric square matrix. Then an n$\times n$ matrix**
$\frac{\partial y}{\partial X} = \mathbf{2aa}\ ' - diag(aa')$.

**where diag(.) denotes the diagonal matrix based on the diagonal elements of the indicated matrix expression.**

## 3 Differentiation of the trace of a matrix

**Convention. If it is desired to differentiate, say, tr(AB) with respect to the elements of A, the operation involved will be interpreted as the "rematricization" of the vector in the following sense**

**Note that $\frac{\partial tr(AB)}{\partial vec(A)}$, is a vector. (exercise: what dimension ?).**

**Having evaluated it we put the resulting vector in a matrix form expressed by $\frac{\partial tr(AB)}{\partial A}$.**

**Let A be a square matrix of order m. Then $\frac{\partial tr(A)}{\partial A} = \mathbf{I}$.**

**If the elements of A are functions of the r-element vector $\theta$, then**

$\frac{\partial tr(A)}{\partial\theta} = \frac{\partial tr(A)}{\partial vec(A)} \frac{\partial vec(A)}{\partial\theta} = \mathbf{vec(I)}' \frac{\partial vec(A)}{\partial\theta}$

**Differentiation of the Trace of a product of a number of matrices. Differentiating the trace of products of matrices with respect to the elements of one of the matrix factors is a special case of differentiation of (1) Linear form: a$'x$, where a is a vector, (2) Nonlinear forms including a bilinear form: z$'Ax$, where A a matrix, and z and x appropriately dimensioned vectors. and quadratic form x$'Ax$. Hence the second-order partials are easily follow from the corresponding results regarding these forms.**

**(1) Trace of a Linear Forms: Derivative of tr(A X )=** $A'$ :

**Let A be T**$\times n$, **and X be n**$\times T$; **then** $\frac{\partial tr(AX)}{\partial X} = \mathbf{A}'$.

**Exercise: Verify that** $\frac{\partial tr(A'B)}{\partial B} = \mathbf{A}$
**Exercise: Verify this by first choosing A**$= I$, **the identity matrix.**
**If X is a function of the elements of the vector** $\theta$, **then**

$\frac{\partial tr(AX)}{\partial\theta} = \frac{\partial tr(AX)}{\partial vec(X)} \frac{\partial vec(X)}{\partial\theta} = \mathbf{vec(A')}' \frac{\partial vec(X)}{\partial\theta}$

**(2) Trace of Nonlinear Forms:**
**Let A be T**$\times n$, **X be n**$\times q$, **and B be q**$\times q$; **then** $\frac{\partial}{\partial X}$ **tr(AXB)** $= \mathbf{A}'B'$.

**If x is a function of the r-element vector** $\theta$ **then** $\frac{\partial}{\partial\theta}$ **tr(AXB)** $= \mathbf{vec(A}'B)' \frac{\partial X}{\partial\theta}$.

**Derivatives of the trace of four matrices, skip the w.r.t matrix and prime them all. Let A be T**$\times n$, **X be n**$\times q$, **B be q**$\times r$, **and Z be r**$\times q$; **then**

$\frac{\partial tr(AXBZ)}{\partial X} = \mathbf{A}'Z'B'$, **and** $\frac{\partial tr(AXBZ)}{\partial Z} = \mathbf{B}'X'A'$.
**If X and Z are functions of the r-element vector** $\theta$, **then**

$\frac{\partial tr(AXBZ)}{\partial \theta}$ = **vec(A**$'Z'B')'$ $\frac{\partial vec(X)}{\partial \theta}$ + **vec(B**$'X'A')'$ $\frac{\partial vec(Z)}{\partial \theta}$

**Let A be T**$\times T$, **X be q**$\times T$, **B be q**$\times q$; **then**

$\frac{\partial tr(AX'BX)}{\partial X}$ = **B**$'XA'$ + **BXA.**

**Exercise: Verify that** $\frac{\partial tr(X'AXB)}{\partial X}$ = **AXB** + **A**$'XB'$
**If X is a function of the r-element vector** $\theta$, **then**

$\frac{\partial tr(AX'BX)}{\partial y}$ = **vec(X)**$'[(A'$ ⊗ **B)** + **(A** ⊗ **B**$')]$ $\frac{\partial vec(X)}{\partial y}$.

**Derivative of a power of trace is that power times the inverse:** $\frac{\partial tr(A^n)}{\partial A}$ = **n**$A^{-1}$

# 4 Differentiation of Determinants.

**Let A be a square matrix of order T; then**

$\frac{\partial |A|}{\partial A} = A^*$, **where** $A^*$ **is the matrix of cofactors of A defined as C**$= (c_{ij})$, **where i,j the element is**

$c_{ij} = (-1)^{i+j} det(A_{ij})$ **where** $A_{ij}$ **denotes an (T-1)**$\times(T$**-1) matrix obtained by deleting i-th row and j-th column ofA.**
**If the elements of A are functions of the r elements of the vector** $\theta$, **then**

$\frac{\partial |A|}{\partial \theta}$ = **vec(**$A^*)'$ $\frac{\partial vec(A)}{\partial \theta}$.
**Derivative of the log of a determinant is simply transpose of its inverse:**

$\frac{\partial log(detA)}{dA} = (A^{-1})'$

**If A is symmetric,** $\frac{\partial log(detA)}{dA}$ = **2**$A^{-1} - diag(A^{-1})$ **( Note,** $a_{ij} = a_{ji}$)

where diag(W) denotes a matrix based on the selection of diagonal terms only of W.

## 5 Further Matrix Derivative Formulas for (Long) Vec and Inverse Matrices

Derivative of a product of three matrices with respect to the vec operator applied to the middle matrix involves a kronecker product of the transpose of the last and the first matrix.
$\frac{\partial vec(AXB)}{\partial vecX} = \mathbf{B}' \otimes \mathbf{A}$
Derivative of the inverse matrix with respect to the elements of the original matrix

$\frac{\partial (X^{-1})}{\partial x_{ij}} = -X^{-1}ZX^{-1}$

where Z is a matrix of mostly zeroes except for a 1 in the (i,j)$^{th}$ position. In this formula if $x_{ij} = x_{ji}$ making it a symmetric matrix, the same formula holds except that the matrix Z has one more 1 in the symmetric (j,i)$^{th}$ position. Why minus? When the matrix X is 1×1 or scalar we know from elementar calculus that the derivative of (1/X) is $-1/X^2$. The above formula is a generalization of this.
Exercise: Verify the above formulas with simple examples.

## 6 Some Matrix Results for Multivariate Normal Variables

If x $\sim$ N($\mu$, V), x is an n dimensional (multivariate) normal with mean vector $\mu$ and the n$\times n$ covariance matrix V, a linear transformation of x is also normal. Consider the linear transformation y = A x + b, where y and b are n×1 vectors and A is an $n \times n$ matrix. Thus
x$\sim$ N($\mu$, V) and y = A x + b $\Rightarrow$ y $\sim$ N(b + A$\mu$, A V A$'$ ) (1)

If x and y are jointly multivariate normal random variables

$$\left[ \begin{smallmatrix} x \\ y \end{smallmatrix} \right] \sim \mathbf{N}\left( \left[ \begin{smallmatrix} E(x) \\ E(y) \end{smallmatrix} \right], \begin{bmatrix} V_{xx} & V_{xy} \\ V_{yx} & V_{yy} \end{bmatrix} \right) \quad (1)$$

**then the conditional distribution of x conditional on y is also normal**

$$\mathbf{x}|\ \mathbf{y} \sim \mathbf{N}(\ [\ \mathbf{E(x)} + V_{xy}V_{yy}^{-1}\ (\mathbf{y}-E(y)]\ ,\ V_{xx} - V_{xy}V_{yy}^{-1}V_{yx}\ ) \qquad \mathbf{(2)}$$

**Similarly, y conditional on x is also normal.**

$$\mathbf{y}|\ \mathbf{x} \sim N\ (\ [\ \mathbf{E(y)} + V_{yx}V_{xx}^{-1}\ (\mathbf{x}-E(x)]\ ,\ V_{yy} - V_{yx}V_{xx}^{-1}V_{xy}\ ) \qquad \mathbf{(3)}$$

**If x is a $p \times 1$ vector of multivariate normal variables with mean vector $\mu$ and covariance matrix V, that is, x $\sim$ N$(\mu, V)$, and let A and B be $p \times p$ matrices of constants, it can be shown that**

$\mathbf{E}[(x'Ax)(x'Bx) = tr(AV)tr(BV) + 2\ tr(AVBV) + (\mu'A\mu)tr(BV) + (\mu'B\mu)tr(AV) +$
$+\ \mathbf{4}\mu'(AVB)\mu + (\mu'A\mu)(\mu'B\mu)$
$\mathbf{Cov}[(x'Ax)(x'Bx)] = 2tr(AVBV) + 4\mu'(AVB)\mu.$
$\mathbf{var(x'}Ax) = 2tr(AV)^2\ +\ \mathbf{4}\mu'(AVB)\mu$

**For a proof see Graybill(1983, Matrices with Applications in Statistics, Wadsworth, Belmont, Calif., p367)**

# 7 Taylor Series in Matrix Notation

**In calculus Taylor's Theorem is described as follows. We use the operator notation:**

$(\mathbf{h}\ \frac{\partial}{\partial x} + \mathbf{k}\ \frac{\partial}{\partial y}\ )\ \mathbf{f}(x_o,\ y_o) = \mathbf{h}\ f_x(x_o,\ y_o)\ +\ \mathbf{k}\ f_y(x_o,\ y_o)$

**where $f_x$ denotes $\partial f/\partial x$ and $f_y$ denotes $\partial f/\partial y$. Similarly, for the 2-nd power:**

$$(\mathbf{h}\ \frac{\partial}{\partial x} + \mathbf{k}\ \frac{\partial}{\partial y}\ )^2\ \mathbf{f}(x_o,\ y_o) = h^2\ f_{xx}(x_o,\ y_o)\ + 2hkf_{xy}(x_o,\ y_o)\ +\ k^2\ f_{yy}(x_o,\ y_o)$$

**where $f_{xx} = \partial^2 f/\partial x \partial x,\ f_{xy} = \partial^2 f/\partial x \partial y$ and $f_{yy}$ denotes $\partial^2 f/\partial y \partial y$. In general,**

(h $\frac{\partial}{\partial x}$+ k $\frac{\partial}{\partial y}$ $)^n$ f$(x_o, y_o)$ is obtained by evaluating a Binomial expansion of the n-th power.

Given that the first n derivatives of a function f(x,y) exist in a closed region, and that the (n+1)$-st$ derivative exists in an open region. Taylor's Theorem states that

$$\text{f}(x_o + h, y_o + k) = f(x_o, y_o) + (h \frac{\partial}{\partial x} + \text{k} \frac{\partial}{\partial y}) \text{ f}(x_o, y_o) + \frac{1}{2!} (\text{h} \frac{\partial}{\partial x} + \text{k} \frac{\partial}{\partial y})^2 \text{ f}(x_o, y_o) + \\ \cdots +$$

$$+ \frac{1}{n!}(h \frac{\partial}{\partial x} + \text{k} \frac{\partial}{\partial y})^n \text{ f}(x_o, y_o) + \frac{1}{(n+1)!}(h \frac{\partial}{\partial x} + \text{k} \frac{\partial}{\partial y})^{n+1} \text{ f}(x_o + \alpha h, y_o + \alpha k)$$

where 0$< \alpha < 1$, x$= x_o + h$, y$= y_o + k$. The term involving the (n+1)$-th$ partial is called the remainder term. When one writes this in the matrix notaion it is usually intended as an approximation, and all terms containing higher than second order partials are ignored. Let x be a p$\times$1 vector with elements $x_1$, $x_2, \cdots, x_p$. Now f(x) is a function of p variables. Let $\partial f(x)/\partial x$ denote a p$\times$1 vector with elements $\partial f(x)/\partial x_i$ , where i$= 1, 2, \cdots p$. Similarly,
let $\partial^2 f(x)/\partial x \partial x'$ denote a p$\times p$ matrix with (i,j)-th element $\partial^2 f(x)/\partial x_i \partial x_j$.
Now Taylor's approximation is

$$\text{f(x)} = f(x^o) + \Sigma_{i=1}^p \ (x_i - x_i^o) \ \frac{\partial}{\partial x} f(x^o) \ + \frac{1}{2} \Sigma_{i=1}^p \Sigma_{j=1}^p \ (x_i - x_i^o) [\ \frac{\partial^2}{\partial x_i \partial x_j} f(x^o)] \ (x_j - x_j^o),$$

that is,

$$\text{f(x)} = f(x^o) + (\text{x} - x^o)' \ \frac{\partial f(x^o)}{\partial x} \ + \frac{1}{2} \ (\text{x} - x^o)' [\ \frac{\partial^2 f(x^o)}{\partial x \partial x'}] \ (\text{x} - x^o)$$

in the matrix notaion. If the second derivative term is to be the remainder term we replace $\partial^2 f(x^o)$ by $\partial^2 f(\bar{\mathbf{x}})$ where $\bar{\mathbf{x}}$= $\alpha$ $x^o$+ (1$-\alpha$) x, with 0$\leq \alpha \leq 1$.

## 8 Matrix Inverse by Recursion

If we want to compute the inverse of (A$-zI)^{-1}$ recursively let us write it as

$(\mathbf{A} - zI)^{-1} = -\ (\mathbf{zI} - A)^{-1} = \frac{1}{\triangle}\ [\ \mathbf{I} z^{n-1} + B_1 z^{n-1} + \cdots + B_{n-1}]$

where

$\triangle$ = det (zI$-A$) $= z^n + a_{n-1}\ z^{n-1} + \cdots + a_0$

is a polynomial in complex number z of the so-called z-transform. This is often interpreted as z= $L^{-1}$, where L is the lag operator in time series ($\mathbf{L}x_t = x_{t-1}$). Multiplying both sides by $\triangle$ we get polynomials in z on both sides. Equating the like powers of z we obtain the recursion:

$B_1 = \mathbf{A} + a_{n-1}\ \mathbf{I}$

$B_2 = \mathbf{A}B_1 + a_{n-2}\ \mathbf{I}$

$B_k = \mathbf{A}B_k + a_{n-k}$ **I for k = 2, $\cdots$, n-1**

and finally since $B_n$ is absent above, we have

$B_n = \mathbf{0} = \mathbf{A}B_{n-1} + a_0\ \mathbf{I}$

# 9 Matrix Inversion When Two Terms Are Involved

Let G be an n $\times$ n matrix defined by

$$\mathbf{G} = [\mathbf{A} + \mathbf{BDB'}]^{-1}, \tag{1}$$

where A and D are nonsingular matrices of order n and m respectively, and B is n $\times$ m. Then

$$\mathbf{G} = A^{-1} - A^{-1}\ \mathbf{B}\ [D^{-1} + \mathbf{B}'A^{-1}B]^{-1}B'A^{-1}. \tag{2}$$

The result is verified directly by showing that $\mathbf{G}G^{-1} = \mathbf{I}$. Let E= $D^{-1} + \mathbf{B}'A^{-1}B$. Then

$G^{-1}G = \mathbf{I} - \mathbf{B}E^{-1}B'A^{-1} + \mathbf{BDB}'A^{-1} - \mathbf{BDB}'A^{-1}BE^{-1}B'A^{-1}$

$= \mathbf{I} + [-BE^{-1} + \mathbf{BD} - \mathbf{BDB}'A^{-1}BE^{-1}]B'A^{-1}$

$= \mathbf{I} + \mathbf{BD}[-D^{-1}E^{-1} + \mathbf{I} - \mathbf{B}'A^{-1}BE^{-1}]B'A^{-1}$

$= \mathbf{I} + \mathbf{BD}[\mathbf{I}-EE^{-1}]B'A^{-1} = \mathbf{I}$, where $(\mathbf{I}-I) = \mathbf{0}$ eliminates the second term.

An important special case arises when B= $b$ is an n $\times$ 1 vector, and D = 1. Then

$$(\mathbf{A} + \mathbf{bb}')^{-1} = A^{-1} - \frac{A^{-1}bb'A^{-1}}{1+b'A^{-1}b.} \tag{3}$$

Many other useful results on matrix inversion are found in Jazwinski (1970, pp. 261-2).

## 10 Useful Results Normal / Chi-square Distributions

1. **If y is normal, N($\mu, \sigma^2$), then z = b + ay is normal, N(a$\mu$ + b, $a^2\sigma^2$) for a $\neq$ 0.**

2. **If y is N($\mu, \sigma^2$), then z = (y $-$ $\mu$)/$\sigma$ is N(0, 1).**

3. **If y and z are independent normal variables, then y + z is normal.**

4. **If each $y_i$, where i = 1, 2, /$\cdots, n$, is independent N($\mu, \sigma^2$), then defining $\bar{\mathbf{y}}$ = $\Sigma_i y_i/n$ implies**

   **$\bar{\mathbf{y}}$ is N($\mu, \sigma^2/n$).**

5. **If y is a random variable with mean $\mu$ and variance $\sigma^2$, then for n sufficiently large, then mean**

   **$\bar{\mathbf{y}}$ from a random sample of size n, is approximately normal N($\mu$, $\sigma^2/n$).**

### 10.1 Chi-square Distributions

6. **If y is N(0,1), then $y^2$ is $\chi^2_1$.**

7. **If $z_1$ is $\chi^2_n$ and $z_2$ is $\chi^2_m$ and $z_1$ and $z_2$ are independent, then $z_1$ + $z_2$ is $\chi^2_{m+n}$.**

8. **If $y_1$, $y_2, \cdots$, $y_n$ are each independent N(0,1), then $\Sigma_i y_i^2$ is distributed as $\chi^2_n$.**

9. **If $z_1$, $z_2, \cdots$, $z_n$ are each independent N($\mu$, $\sigma^2$), then $\Sigma_i$ $(z_i - \mu)^2/\sigma^2$ is distributed as $\chi^2_n$.**

10. **If the sample variance of a random sample $y_1$, $y_2$, $\cdots$, $y_n$ is $\hat{\sigma}^2 = \sum_i^{\infty} (y_i -$ $\bar{\mathbf{y}})^2/(n-1)$, and each $y_i$ is N($\mu, \sigma^2$), then (n$-$1)$\hat{\sigma}^2/\sigma^2$ is $\chi^2_{n-1}$.**

11. **If the (n$\times$1) vector y is distributed N(0, $I_n$), then y $'$y is distributed as $\chi^2_n$.**

12. **If y is an (n $\times$ 1 ) vector distributed as N(0,I$_n$) and A is an ($n\times n$) symmetric idempotent matrix of rank r, then y $'A$ y is distributed as $\chi^2_r$.**

13. **If the (n$\times$1) vector z is distributed as N(0,$\sigma^2 I_n$) and A is an (n$\times n$) symmetric idempotent matrix of rank r, then (z$'A$z/$\sigma^2$) is distributed as $\chi^2_r$.**

14. **Let y be an (m$\times$1) vector that is distributed as N($\delta, \sigma^2 I_n$), A an (n$\times n$) symmetric idempotent matrix such that A$\delta$ = 0, B an (m$\times n$) matrix and AB$'$ = 0. Then B y is distributed independently of the quadratic form (y $'A$y) .**

15. **If y is an (n$\times$1) vector that is distributed as N(0, $\sigma^2 I_n$) and A and B are idempotent (n$\times n$) matrices of rank r and s and AB = 0, then y$'Ay$ is distributed independently of the quadratic form y$'By$.**

16. **If y is an (n$\times$1) vector distributed as N(0, $\sigma^2 I_n$) and A and B are idempotent (n$\times n$) matrices of rank r and s respectively, then u, the ratio of y $'A$y$/\sigma^2$ and y$'By/\sigma^2$ each divided by its rank is distributed as $F_{r,s}$.**

17. **If y is $F_{n,m}$ , then z = 1/y is distributed as $F_{m,n}$.**

18. **If the n$\times$1 vector e is distributed as N($\mu, A$), then e$'A^{-1}e$ is distributed as $\chi^2_{m,\gamma}$ with $\gamma = 0.5\mu' A^{-1}\ \mu$.**

19. **If the (n$\times$1) vector z is distributed as N($\mu, \sigma^2 I$), then W = z$'Az/\sigma^2$, where A is a symmetric idempotent matrix of rank k, is distributed as $y^2_{k,\gamma}$ with noncentrality parameter $\gamma = \mu/2\sigma^2$. If B is an (n$\times n$) symmetric idempotent matrix of rank q with BA = 0, B$\mu$ = 0 and Z = z$'Bz/\sigma^2$, then the quantity u = (W/Z)(q/k) is distributed as non-central $F_{k,q,\gamma}$.**

20. **If the random variable z has a $\chi^2_{k\gamma}$ distribution that is independent of w, which is distributed as $\chi^2_q$, then u = qz/(kw) is distributed as $F_{k,q,\gamma}$.**

21. **Let the (k$\times$1) vector $\hat{\beta}$ be distributed as N($\beta, \sigma^2(X'X)^{-1}$) and $\hat{\beta}_i$ is distributed as N($\beta_i, \sigma^2 c_{ii}$), where $c_{ii}$ is the ith diagonal element of $(X'X)^{-1}$ = C. Consequently $(\hat{\beta}_i - \beta_i)/\sigma c_{ii}$ is distributed as N(0,1) and it is independent of (T$-k)\hat{\sigma}^2/\sigma^2$, which is distributed as $\chi^2_{T-k}$. Therefore,**
    **v** $= \frac{\hat{\beta}_i - \beta_i}{\sigma c_{ii}} \frac{(\sigma^2)^{0.5}}{(\hat{\sigma}^2)^{0.5}} = \frac{\beta_i - \beta}{\hat{\sigma} c_{ii}}$
    **is distributed as Student's $t_{T-k}$.**

22. **If the (n $\times$1) random vector z is N(0,$\sigma^2 I_n$) the expected value of the quadratic form z$'Az/\sigma^2$ is equal to trA. Therefore, if A is an (n$\times n$) symmetric idempotent matrix of rank r, then E(z$'A$z)$\sigma^2$ = r.**

23. **The reciprocal of the central $\chi^2$ random variable with degrees of freedom has expected value E$[(\chi^2_r)^{-1}]$ = 1/(r$-2$), for $r \geq 3$.**

24. **The central $\chi^2$ random variable with r degrees of freedom has variance 2r.**

25. **The square of the reciprocal of a central $\chi^2$ random variable with r degrees of freedom has expected value E$[(\chi^2_r)^{-2}]$ = 1/[(r$-2)(r-4)]$.**

26. **Any non-central $\chi^2$ random variable with r degrees of freedom and noncentrality parameter $\gamma$, may be represented as a central $\chi^2$ random variable**

with (r + 2j) degrees of freedom (conditional on j) where j is a Poisson random variable with parameter $\gamma$.

# 11 Matrix Algebra Review for Normal Theory in Statistics

1. The characteristic roots of an ($n \times n$) matrix A are the n roots of the polynomial $|A - \gamma I|$, where $\gamma$ is a scalar.

2. For an ($n \times n$) orthogonal matrix C (where $\mathbf{C}'C = \mathbf{CC}' = \mathbf{I}$) $\mathbf{C}'$ is orthogonal.

3. If C is orthogonal then its determinant $|C|$ is either 1 or $-1$.

4. If A is an ($n \times n$) symmetric matrix, then there exists an ($n \times n$) matrix P which is orthogonal and $\mathbf{P}'AP =$ a diagonal matrix with diagonal elements that are the characteristic roots of A and the rank of A is equal to the number of non-zero roots.

5. If A is an ($n \times n$) symmetric matrix then A is positive definite if and only if all its characteristic roots are positive, where a positive definite matrix is one where the quadratic form $\mathbf{y}'A\mathbf{y}$ is positive for all $\mathbf{y} \neq 0$.

6. If A is an ($n \times n$) positive definite matrix, then $|A| > 0$, the rank of A is equal to n and A is non-singular.

7. If A is an ($n \times n$) positive definite matrix and P is an ($n \times m$) matrix with rank m, then $\mathbf{P}'AP$ is positive definite.

8. If A is an ($n \times n$) positive definite matrix then there exists a positive definite matrix $A^{-1/2}$ such that $A^{-1/2}AA^{-1/2} = \mathbf{I}$ and $A^{-1/2}A^{-1/2} = A^{-1}$. Also we write $A^{1/2} = (A^{-1/2})^{-1}$ and $A^{1/2}A^{1/2} = \mathbf{A}$.

   In particular, if P is an orthogonal matrix such that $\mathbf{P}\begin{bmatrix} \gamma_1 & 0 \\ 0 & \gamma_p \end{bmatrix}\mathbf{P}' = \mathbf{A}$, where the $\gamma_i$ are the characteristic roots of A, then $A^{1/2} = \mathbf{P}\,[\,\gamma_1^{1/2} \quad .. \quad \gamma_p^{1/2}\,]$ $\mathbf{P}'$.

9. Given an ($n \times n$) symmetric idempotent matrix A (i.e., $\mathbf{A} = \mathbf{A}'$ and $\mathbf{AA} = \mathbf{A}$), then if A is of rank r, A has r characteristic roots equal to 1 and ($n - r$) roots equal to zero, the rank of A is equal to trA and there is an orthogonal matrix C such that

   $$\mathbf{C}'AC = \begin{bmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{bmatrix}.$$

10. **The identity matrix is the only non-singular idempotent matrix and a symmetric idempotent matrix is positive semi-definite.**

11. **If A and B are (n$\times n$) symmetric matrices and B is positive definite, there exists a non-singular matrix Q such that Q$'AQ = \Lambda$ and Q$'BQ =$ I, where $\Lambda$ is a diagonal matrix [Rao (1973, p. 41)].**

12. **If A and B are two symmetric matrices, a necessary and sufficient condition for an othogonal matrix C to exist such that C$'AC = \Lambda$ and C$'BC =$ M, where $\Lambda$ and M are diagonal, is that A and B commute, i.e., AB = BA.**

13. **If A is a symmetric (n$\times n$) matrix with characteristic roots $\lambda_1, \lambda_2, \cdots, \lambda_n$ and corresponding characteristic vectors $p_1, p_2, \cdots, p_n$, then A $= \lambda_1 p_1 p_1' + \cdots + \lambda_n p_n p_n', I = p_1 p_1' + p_2 p_2' + \cdots + p_n p_n'$ and su$p_y$ (y$'A$y/y$'$y) $= \lambda_1$; in$f_y$ (y$'A$y/y$'$y) $= \lambda_n$, where y is a column vector.**

14. **The characteristic roots of A are those of BA$B^{-1}$, where A,B are non-singular matrices.**

15. **If B is (n$\times n$) non-singular matrix and $\eta$ is (n$\times 1$) column vector, then ma$x_y$ (y$'\eta\eta'$y/y$'B$y) $= \eta\ 'B^{-1}\eta$.**