Study of elementary calculus extended to square matrices

Norman C. Davis

The University of Montana

Follow this and additional works at: https://scholarworks.umt.edu/etd

Let us know how access to this document benefits you.

Recommended Citation
Davis, Norman C., "Study of elementary calculus extended to square matrices" (1952). Graduate Student Theses, Dissertations, & Professional Papers. 8245.
https://scholarworks.umt.edu/etd/8245
A STUDY OF ELEMENTARY CALCULUS
EXTENDED TO SQUARE MATRICES

by

Norman C. Davis
B.A., Montana State University, 1950

Presented in partial fulfillment
of the requirements for the degree of
Master of Arts

MONTANA STATE UNIVERSITY
1952
This thesis has been approved by the Board of Examiners in partial fulfillment of the requirements for the degree of Master of Arts.

Chairman of the Board of Examiners

Dean of the Graduate School

Date: Aug 18, 1952
ACKNOWLEDGEMENTS

The author is especially indebted to Professor T. G. Ostrom for his generous assistance and constant guidance throughout the work on this thesis. To Professor G. Marsaglia he is indebted for the suggestion of the subject of the thesis.

N. C. D.
# Table of Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I. Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>II. Algebraic Considerations</strong></td>
<td>3</td>
</tr>
<tr>
<td><strong>III. Series in Matrices</strong></td>
<td>11</td>
</tr>
<tr>
<td><strong>IV. Differentiation</strong></td>
<td>15</td>
</tr>
<tr>
<td><strong>V. Power Series</strong></td>
<td>21</td>
</tr>
<tr>
<td><strong>VI. Matrices of Differential Operators</strong></td>
<td>29</td>
</tr>
<tr>
<td><strong>VII. Integration</strong></td>
<td>33</td>
</tr>
<tr>
<td><strong>VIII. Applications in Mathematical Statistics</strong></td>
<td>35</td>
</tr>
<tr>
<td><strong>Bibliography</strong></td>
<td>38</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

The purpose of this paper is to make a study of square matrices and their manipulation in order to extend certain rules, definitions, and theorems of the calculus of scalar variables to functions of matrices. Unless otherwise stated, all matrices to be considered will be square.

Matrix theory is of particular importance to all the fields of applied mathematics since many involved problems can be stated simply and solved more easily by expressing them as matrix functions. The topics discussed in this paper will, in general be those that are useful in mathematical statistics.

Certain algebraic considerations will be introduced and used to clarify other more complicated ideas as well as to prove certain theorems.

An understanding of the calculus of scalar variables, as well as, an understanding of the more elementary rules of matrix manipulation will be pre-supposed on the part of the reader.

Conventional notation will be used in most cases, however, the notation is listed below because it is especially important that it be understood by the reader.
1. \((a_{rs})\) denotes a matrix of order \(n\) with elements \(a_{rs}\) where \(r,s = 1,2,\ldots,n\).

2. \(|A|\) denotes the determinant of the matrix \(A\).

3. \(\{x \ y\} \) will denote the sum of \(s\) \(x\)'s and \(t\) \(y\)'s multiplied together in every possible way.

4. \(A'\) denotes the transpose of the matrix \(A\).

5. \(A^{-1}\) denotes the inverse of the matrix \(A\).

6. "\(A\) is an \(nxm\) matrix" means that the matrix \(A\) has \(n\) rows and \(m\) columns.

7. If \(A_1, A_2, \ldots, A_n\) is a set of matrices, \(a_{irs}\) denotes the element in the \(r\)th row and \(s\)th column of the matrix \(A_i\) and we write \(A_i = (a_{irs})\).

8. \(I_n\) denotes the identity matrix of order \(n\).
In this chapter, we shall introduce certain algebraic forms and note their expression as matrices.

**Definition:** If a function \( A(y, x) \) is linear and homogenous in the variables of two sets of \( n \) variables, \( y, x \), it is called a bilinear form. Such a function has \( n^2 \) coefficients and may be written:

\[
A(y, x) = y_1(a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n)
\]

Let \( x \) be the column matrix, \((x_1, x_2, \ldots, x_n)\) and let \( y \) be the row matrix, \((y_1, y_2, \ldots, y_n)\). Then:

\[
A(y, x) = y(a_{rs})x\text{ where } (a_{rs})\text{ has order } n.
\]

If we let \((a_{rs}) = a\), then:

\[
A(y, x) = yax = x^T ax
\]

**Definition:** When the sets of variables in a bilinear form are identical so that \( y = x \), the bilinear form is called a quadratic form. The quadratic form is a homogenous function of the second degree.

The coefficient of \( x_i x_j \), \((i \neq j)\), is \( a_{ij} + a_{ji} \). We note that this coefficient will be unaltered if \( a_{ij} \) and \( a_{ji} \) are both changed to \( \frac{1}{2}(a_{ij} + a_{ji}) \). The quadratic form, then, is:

\[
A(x, x) = x^T ax
\]

where \( a \) is a symmetric matrix and \( x = (x_1, x_2, \ldots, x_n) \).
Consider the linear homogeneous function of the matrices \( A_1, A_2, \ldots, A_n \):

\[
P(A_1, A_2, \ldots, A_n) = \sum_{i=0}^{n} \alpha_i A_i, \quad i = 0, 1, 2, \ldots, n,
\]
where \( \alpha_i \) is a scalar. From the properties of addition of matrices, then, \( P(A_1, A_2, \ldots, A_n) \) is a matrix,

\[
A = (a_{rs})
\]
such that

\[
a_{rs} = \sum_{i=0}^{n} \alpha_i a_{irs}.
\]

Thus, if

\[
A_i = A^i,
\]

\[
P(A) = \alpha_1 A + \alpha_2 A^2 + \ldots + \alpha_n A^n + \alpha_0 I_n.
\]

**Definition:** If \( A \) is a matrix of order \( n \) and \( n \) is any positive integer, \( A^m \) is the product \( AAA \ldots \) to \( m \) factors.

**Definition:** \( A^0 \equiv I_n \).

**Theorem:** \( A^m A^k = A^{m+k} \).

**Theorem:** \( (A^m)^k = A^{mk} = A^{km} = (A^k)^m \).

The two theorems above follow from the associative law of multiplication.

**Definition:** Let \( A \) be a non-singular matrix, so that \( A^{-1} \) exists. Then, the negative integral powers of \( A \) are defined to be:

\[
A^{-m} \equiv (A^{-1})^m.
\]

Thus, in the multiplication of matrices, the usual laws of scalar algebra hold for integral exponents.

**Definition:** If \( A \) is a matrix such that:
where $B$ is a matrix, we may write:

$$\frac{1}{B^m} = A$$

and we call $A$ the $m^{th}$ root of $B$.

**Definition:** Let $u$ be a matrix of order $n$ and $p_i$ be scalar constants, $i = 0, 1, 2, \ldots, m$. Then,

$$P(u) = p_0 u^m + p_1 u^{m-1} + p_2 u^{m-2} + \ldots + p_{m-1} u + p_m I_n$$

is called a polynomial of the matrix $u$.

The scalar polynomial of $x$ corresponding to $P(u)$ is

$$P(x) = p_0 x^m + p_1 x^{m-1} + \ldots + p_{m-1} x + p_m.$$

Matrices obey the associative law and are commutative with scalars. We have shown that powers of $u$ are commutative with each other. Therefore, the multiplication and addition of matric polynomials is isomorphic to the multiplication and addition of the corresponding scalar polynomials. Thus, the identity in scalars,

$$P_1(x)P_2(x) \equiv P_3(x)$$

implies a corresponding matric identity,

$$P_1(u)P_2(u) \equiv P_3(u).$$

It follows, then, that, if $a_1, a_2, \ldots, a_m$ are the roots of

$$P(x) = 0,$$

$$P(x) = p_0(x - a_1)(x - a_2) \ldots (x - a_m),$$

and the corresponding matric polynomial may be written in the factorized form:
Example:
Let \( u = \begin{pmatrix} 3 & 1 \\ 4 & 1 \end{pmatrix} \) and suppose \( P(u) = 2u^2 - 10u + 12I \).

Then:

\[
P(u) = 2 \begin{pmatrix} 13 & 4 \\ 16 & 5 \end{pmatrix} - 10 \begin{pmatrix} 3 & 1 \\ 4 & 1 \end{pmatrix} + 12 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}
\]

\[
= \begin{pmatrix} 8 & -2 \\ -8 & 12 \end{pmatrix}
\]

In this case:

\[
P(x) = 2x^2 + 10x + 12 = 2(x - 2)(x - 3).
\]

Hence:

\[
P(u) = 2(u - 2I)(u - 3I) = 2 \begin{pmatrix} 1 & 1 \\ 4 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 4 & -2 \end{pmatrix}
\]

\[
= \begin{pmatrix} 8 & -2 \\ -8 & 12 \end{pmatrix}
\]

Theorem: Let \( P(u) \) be the polynomial:

\[
P(u) = p_0u^m + p_1u^{m-1} + \ldots + p_{m-1}u + p_mI
\]

and let \( b \) be any constant. Then, \( bi-u \) is a factor of \( P(bI) - P(u) \).

Proof: Let \( P(x) = p_0x^m + p_1x^{m-1} + \ldots + p_m \).

Then,

\[
P(b) - P(x) = \sum_{r=0}^{m} p_r (b^{m-r} - x^{m-r})
\]

The expression on the right is divisible by \( (b - x) \), so that:

\[
P(b) - P(x) = (b-x)S(b,x)
\]

where \( S(b,x) \) is a function of degree \( m-1 \) in both \( x \) and \( b \). By the discussion on the preceding page, it follows that:

\[
P(bI) - P(u) = (bI - u)S(bI,u)
\]

Lagrange's Interpolation Formula: Let \( a_1, a_2, \ldots, a_n \).
be distinct arbitrary constants, then, if the degree of \( P(x) \)
does not exceed \( n-1 \), it is well known for scalar variables,
that:

\[
P(x) = \sum_{r=1}^{n} P(a_r) \frac{s^{\tau}(as - x)}{s^{\tau}(as - ar)}
\]

The above scalar identity leads to the corresponding
matrix identity:

\[
P(u) = \sum_{r=1}^{n} P(a_r) \frac{s^{\tau}(asI - u)}{s^{\tau}(as - ar)}
\]

which holds good for all distinct \( a_r \) provided that the de­
gree of \( P(u) \) does not exceed \( n - 1 \).

We now find it convenient to introduce the idea of
a Lambda-Matrix.

**Definition:** Let \( f \) be a matrix such that:

\[
f = (f_{ij}(\lambda)) = f(\lambda)
\]

where \( f_{ij}(\lambda) \) are rational functions of a scalar parameter.
The matrix \( f \) is then defined to be a \( \lambda \)-matrix.

If \( \lambda \) is the highest degree in \( \lambda \) of any of the ele­
ments, \( f \) is said to be of degree \( \lambda \). Such a matrix can evid­
ently be expanded in the form:

\[
f = A_0 \lambda^N + A_1 \lambda^{N-1} + \ldots + A_{N-1} \lambda + A_N,
\]

where \( A_0, A_1, \ldots, A_N \) are matrices that are independent of
\( \lambda \). A \( \lambda \)-matrix is accordingly a polynomial in \( \lambda \) with matrix
coefficients.

**Theorem:** Let \( f \) be a \( \lambda \)-matrix and let \( w \) be a square
matrix of the same order with elements independent of \( \lambda \).
Let:

\[ f_1(w) = w^N A_0 + w^{N-1} A_1 + \ldots + A_N, \]

so that

\[ f(\lambda) - f_1(w) = \sum_{i=0}^{N-1} \left[ (\lambda I)^{N-i} - w^{N-i} \right] A_i. \]

Then, \( f_1(w) \) is the remainder when \( f(\lambda) \) is divided on the left by \( \lambda I - w \). Similarly, \( f_2(w) \) is the remainder when \( f(\lambda) \) is divided on the right by \( \lambda I - w \), where:

\[ f_2(w) = A_0 w^N + A_1 w^{N-1} + \ldots + A_N. \]

**Proof:**

\[ f(\lambda) - f_1(w) = \sum_{i=0}^{N-1} \left[ (\lambda I)^{N-i} - w^{N-i} \right] A_i \]

Now:

\[ (\lambda I)^{N-i} - w^{N-i} = (\lambda I - w)(\lambda I^{N-i-1} + w\lambda^{N-i-2} + \ldots + w^{N-i-1}). \]

Hence \( \lambda I - w \) is a factor on the left of \( f(\lambda) - f_1(w) \), and we may therefore write:

\[ f(\lambda) - f_1(w) = (\lambda I - w)Q(\lambda) \]

where \( Q(\lambda) \) is some \( \lambda \)-matrix. Thus, \( f_1(w) \) is the remainder when \( f(\lambda) \) is divided on the left by \( \lambda I - w \).

By a similar argument, it can easily be shown that, when \( f(\lambda) \) is divided on the right by \( \lambda I - w \), the remainder is:

\[ f_2(w) = A_0 w^N + A_1 w^{N-1} + \ldots + A_N. \]

**Definition:** Let \( w \) be a matrix of order \( n \) and let \( \lambda_i, \ i = 1,2,\ldots,n, \) be the characteristic roots of \( w \). Then, the matrix, \( f(\lambda) \), where:

\[ f(\lambda) = (\lambda I - w)Q(\lambda) \]

is called the characteristic matrix of \( w \). If \( \Delta(\lambda) \) is the
determinant:
\[ \Delta(\lambda) = |\lambda I - w|, \]

\( \Delta(\lambda) \) is called the characteristic function of \( w \) and \( \Delta(\lambda) = 0 \) is called the characteristic equation of \( w \).

**Theorem:** Let \( w \) be a matrix of order \( n \) such that the characteristic roots \( \lambda_1, \lambda_2, \ldots, \lambda_n \) are distinct. Then, any positive integral power of \( w \) is expressible as a linear function of the first \( n-1 \) powers of \( w \).

**Proof:** By the Cayley-Hamilton theorem, \( \Delta(w) = 0 \), but
\[ \Delta(w) = w^n + a_1w^{n-1} + \ldots + a_nI = 0 \]
where \( a_i, i = 1, 2, \ldots, n \) are constants.

Thus:
\[ w^n = -(a_1w^{n-1} + a_2w^{n-2} + \ldots + a_n) \]

**Theorem:** Let \( P(w) \) be any polynomial of \( w \). Then:
\[ P(w) = k_0w^{n-1} + k_1w^{n-2} + \ldots + k_{n-1} \]
where the coefficients \( k_i \) are scalars. This theorem follows as a corollary to the preceding theorem.

If \( \lambda_r \) is a characteristic root of \( w \), it follows that the corresponding scalar polynomial \( P(\lambda_r) \) is reducible to:
\[ P(\lambda_r) = k_0\lambda_r^{n-1} + k_1\lambda_r^{n-2} + \ldots + k_{n-1} \]

**Sylvester's Theorem:** Let \( u \) be an \( nxn \) matrix. If the \( n \) characteristic roots of \( u \) are all distinct and \( P(u) \) is any polynomial of \( u \), then:
\[ P(u) = \sum_{r=1}^{n} P(\lambda_r) \frac{\prod_{j \neq r} (\lambda_j - \lambda_r)}{\prod_{j \neq r} (\lambda_j - \lambda_r)} \]
where \( \lambda_r \) are the characteristic roots of \( u, r = 1, 2, \ldots, n. \)
Proof: By the preceding theorem, $P(u)$ and $P(\lambda)$ are reducible to the similar forms:

$$P(u) = k_0 u^{n-1} + k_1 u^{n-2} + \ldots + k_{n-1}$$
$$P(\lambda) = k_0 \lambda^{n-1} + k_1 \lambda^{n-2} + \ldots + k_{n-1}$$

Now by Lagrange’s Interpolation Formula, with $\lambda$ substituted for $a$, this formula may be used here since $P(u)$ is expressed as an equivalent polynomial of degree not exceeding $n - 1$.

We note that the theorem is also valid if $P(u)$ contains negative integral powers of $u$ ($u$ non-singular).

We shall find this theorem useful in the discussion of power series.
In this chapter, we shall discuss a part of the ideas concerning matrices and reserve a part of the discussion until we have defined a derivative.

Let $A_0, A_1, A_2, \ldots$ be an infinite sequence of matrices of the same order and let:

$$S = A_0 + A_1 + A_2 + \cdots$$

be their sum. Then, from the discussion of the linear function of matrices in Chapter II, it follows that, if

$$S_p = \sum_{i=0}^{p} A_i,$$

then,

$$S_p = \left( \sum_{i=0}^{p} a_{irs} \right)$$

where $A_i = (a_{irs})$.

**Definition:** If every element of $S_p$ converges to a limit as $p$ tends to infinity, then by:

$$S = \lim_{p \to \infty} S_p$$

we shall mean the matrix of limiting elements and we shall say that the infinite series is convergent. If $S$ exists, then, by definition, the matrix infinite series, $\sum A_i$, converges to $S$.

One of the most important cases is that in which the series is of the type:
\[ S_p = \sum_{p=1}^{\infty} k_p A^p \]

where \( A \) is a matrix of order \( n \) and the coefficients \( k \) are all scalar.

**Definition:** We shall define a matric series of the form

\[ S_p = \sum_{n=1}^{p} k_n A^n \]

to be a power series of matrices. It follows from the discussion of \( P(u) \) that \( S_p \) is a matrix whose elements are scalar power series.

**Theorem:** Let \( A \) be a matrix of order \( n \) and let \( U \) be the greatest modulus of any of the elements of \( A \). Then, a sufficient condition for the convergence of:

\[ S = \sum_{p=1}^{\infty} k_p A^p \]

is that the corresponding scalar series in \( nU \) converges.

**Proof:** Let \( A = (a_{rs}) \). No element of \( A \) exceeds \( U \). No element of \( A^2 \) exceeds \( nU^2 \) since:

\[ A^2 = \sum_{s=1}^{n} a_{rs} a_{sr} \], \( r = 1, 2, \ldots, n. \)

Similarly, it follows that no element of \( A^3 \) exceeds \( n^2 U^3 \).

By induction, it follows, that no element of \( A^p \) exceeds \( n^{p-1} U^p \). Thus, every element of \( S_p \) is dominated by the scalar series:

\[ k_0 + k_1 U + nk_2 U^2 + n^2 k_3 U^3 + \ldots \]

\[ = (n-1)k_0 + \frac{1}{n}(k_0 + k_1 \theta + k_2 \theta^2 + \ldots) \]

where \( k_1 \) is the modulus of the corresponding scalar coefficients of the elements of \( S_p \) and \( \theta = nU \).

Thus, the matric power series:

\[ S_p = \sum_{p=1}^{\infty} k_p A^p \]
will converge if
\[
\frac{(n-1)k_0}{n} + \frac{1}{n}(k_0 + k_1\theta + k_2\theta^2 + \ldots)
\]
converges.

**Theorem:** Let \( \lambda_r, r = 1,2,\ldots,n, \) be the characteristic roots of a matrix \( u \) of order \( n \) where the roots are all distinct. Then:

\[
S(u) = \sum_{i=0}^{\infty} k_i u^i,
\]
where \( k_i \) is scalar, converges provided that all the corresponding scalar series \( S(\lambda_r) \) are convergent.

**Proof:** From our discussion in the chapter on algebra of matrices, we note that any polynomial, \( P(u) \), can be expressed as a function of the first \( n-1 \) powers \( u^i \). Thus from Sylvester's Theorem:

\[
S_P(u) = \sum_{i=0}^{P} k_i u^i = \sum_{r=1}^{n} S_P(\lambda_r) z_r
\]
where

\[
z_r = \frac{j \rho(\lambda_j I - u)}{j \rho(\lambda_j - \lambda_r)}.
\]

Then:

\[
S(u) = \lim_{p \to \infty} S_P(u) = \lim_{p \to \infty} \sum_{r=1}^{n} S_P(\lambda_r) z_r
\]

Since the existence of the limit on the right implies the existence of the limit on the left, this theorem is a corollary to Sylvester's Theorem. We shall use this theorem in further discussion of power series of matrices.
Example:

Let \( u = \begin{pmatrix} .15 & -.01 \\ -.25 & .15 \end{pmatrix} \)

and let \( S(u) = I + u + u^2 + u^3 + \ldots \)

Then \( f(\lambda) = \begin{pmatrix} -.15 & .01 \\ .25 & -.15 \end{pmatrix} \)

and \( \Delta(\lambda) = \lambda^2 - .3\lambda + .02 \)

The characteristic roots are \( \lambda_1 = .1 \) and \( \lambda_2 = .2 \)

Then:

\[
\sum_{r=1}^{2} S(\lambda_r) \prod_{j \neq r} \left( \frac{\lambda_j I - u}{\lambda_j - \lambda_r} \right) = \frac{10}{9} \left( \frac{\lambda_2 I - u}{\lambda_2 - \lambda_1} \right) + \frac{10}{8} \left( \frac{\lambda_1 I - u}{\lambda_1 - \lambda_2} \right)
\]

\[
= \frac{10}{9} \begin{pmatrix} .05 & .01 \\ .25 & .05 \end{pmatrix} \frac{1}{1} + \frac{10}{8} \begin{pmatrix} -.05 & .01 \\ .25 & -.05 \end{pmatrix} \frac{1}{1}
\]

\[
= \frac{1}{72} \begin{pmatrix} 85 & -1 \\ -25 & 85 \end{pmatrix} = S(u)
\]
CHAPTER IV

DIFFERENTIATION

Definition: Let \( A(t) \) be a matrix whose elements, \( a_{rs}(t) \), are functions of a variable \( t \). Then, the matrix \( A(t) \) is called a matrix function of \( t \) and may be denoted simply as \( A \).

Definition: We define the derivative of the matrix with respect to the variable \( t \), when such a derivative exists, to be:

\[
\frac{dA}{dt} = \lim_{h \to 0} \frac{A(t+h) - A(t)}{h} = \frac{d}{dt} a_{rs}
\]

where \( h \) is a scalar and \( A(t+h) \) is the matrix whose elements \( a_{rs}(t+h) \) are functions of \( t+h \).

Let:

\[
A = (a_{rs}) \quad \text{and} \quad B = (b_{rs}).
\]

We now consider the fundamental rules of this type of differentiation.

Theorem: \( \frac{d(A+B)}{dt} = \frac{dA}{dt} + \frac{dB}{dt} \)

Proof: \( A + B \) is a matrix with elements \( a_{rs} + b_{rs} \). The theorem then follows from the definition of the derivative.

Theorem: When \( AB \) is defined, i.e. are of the same order,

\[
\frac{d(AB)}{dt} = \frac{dA}{dt}B + A\frac{dB}{dt}
\]
Proof: $AB = (\sum_{i=1}^{n} a_{ri} b_{is})$

$$\frac{d(AB)}{dt} = \frac{d}{dt} \sum_{i=1}^{n} a_{ri}(t)b_{is}(t) = \sum_{i=1}^{n} \frac{d}{dt} a_{ri}(t)b_{is}(t)$$

$$= \sum_{i=1}^{n} \frac{da_{ri}}{dt} b_{is} + \sum_{i=1}^{n} a_{ri} \frac{db_{is}}{dt}$$

and the theorem follows immediately.

**Theorem:** \( \frac{dA'}{dt} = (\frac{dA}{dt})' \)

**Proof:** \( A' \) is a matrix whose elements are \( a_{sr}(t) \).

Then:

$$\frac{dA'}{dt} = \frac{da_{sr}(t)}{dt} = (\frac{dA}{dt})'$$

**Theorem:** When \( A \) is non-singular,

$$\frac{dA^{-1}}{dt} = -A^{-1} \frac{dA}{dt} A^{-1}$$

**Proof:**

$$A A^{-1} = I$$

$$\frac{d(A^{-1} A)}{dt} = A^{-1} \frac{dA}{dt} A^{-1} + A^{-1} \frac{dA}{dt} A^{-1} = 0$$

Therefore:

$$\frac{dA^{-1}}{dt} = -A^{-1} \frac{dA}{dt} A^{-1}$$

**Theorem:**

$$\frac{dA^2}{dt} = \frac{dA A}{dt} + A \frac{dA}{dt}$$

**Proof:** Let \( A = B \) in the theorem on the derivative of product of two matrices.

**Theorem:** In general, when \( m \) is any positive integer,

$$\frac{dA^m}{dt} = \begin{pmatrix} A & \frac{dA}{dt} \\ m-1 & 1 \end{pmatrix}$$
This theorem follows from repeated application of the preceding theorem.

**Theorem:** Let \( A = a_{rs}(t), r, s = 1, 2, \ldots, n. \)

Then:

\[
\frac{d|A|}{dt} = \begin{vmatrix}
\frac{da_{11}(t)}{dt} & a_{12}(t) & \cdots & a_{1n}(t) \\
\frac{da_{21}(t)}{dt} & a_{22}(t) & \cdots & a_{2n}(t) \\
\vdots & \vdots & \ddots & \vdots \\
\frac{da_{n1}(t)}{dt} & a_{n2}(t) & \cdots & a_{nn}(t)
\end{vmatrix}
\]

\[
\frac{d}{dt} \begin{vmatrix}
a_{11}(t) & \frac{da_{12}(t)}{dt} & \cdots & \frac{a_{1n}(t)}{dt} \\
a_{21}(t) & \frac{da_{22}(t)}{dt} & \cdots & \frac{a_{2n}(t)}{dt} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1}(t) & \frac{da_{n2}(t)}{dt} & \cdots & \frac{a_{nn}(t)}{dt}
\end{vmatrix}
\]

\[
\vdots
\]

\[
\frac{d}{dt} \begin{vmatrix}
a_{11}(t) & \frac{da_{12}(t)}{dt} & \cdots & \frac{a_{1(n-1)}(t)}{dt} & \frac{da_{1n}(t)}{dt} \\
a_{21}(t) & \frac{da_{22}(t)}{dt} & \cdots & \frac{a_{2(n-1)}(t)}{dt} & \frac{da_{2n}(t)}{dt} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
a_{n1}(t) & \frac{da_{n2}(t)}{dt} & \cdots & \frac{a_{n(n-1)}(t)}{dt} & \frac{da_{nn}(t)}{dt}
\end{vmatrix}
\]

**Proof:**

\[
\lambda = \sum_{l_1, l_2, \ldots, l_n = 1, 2, \ldots, n, \text{i.e., } l_j \neq i_k, j \neq k}^{a_{l_1, i_2, \ldots, i_n} a_{l_2, i_3, \ldots, i_n} \cdots a_{l_n, i_1, \ldots, i_{n-1}}} \frac{d}{dt} \frac{a_{l_1, i_2, \ldots, i_n} a_{l_2, i_3, \ldots, i_n} \cdots a_{l_n, i_1, \ldots, i_{n-1}}}{n!}
\]

where \( l_1, l_2, \ldots, l_n = 1 \) or \(-1\) according as \( l_1, l_2, \ldots, l_n \) is
an even or an odd permutation of $1,2,3,\ldots,n$.

$$
\frac{d}{dt}l = \frac{d}{dt} \sum_{i=1}^{n} \epsilon_{i1, i2, \ldots, i_n} a_{i1} a_{i2} \cdots a_{i_n}
$$

$$
= \sum_{i=1}^{n} \epsilon_{i1, i2, \ldots, i_n} a_{i1} a_{i2} \cdots a_{i_n}
$$

$$
= \sum_{i=1}^{n} \sum_{k=1}^{n} \epsilon_{i1, i2, \ldots, i_n} a_{i1} a_{i2} \cdots a_{i_n}
$$

Each inner summation is a determinant of the type in the statement of the theorem.

We now define another type of matrix differentiation.

**Definition:** Let $f(A)$ be a function of the matrix $A$ and let $B$ be a non-singular matrix, such that:

$$AB = BA$$

We define $df(A)$ to be:

$$df(A) = \lim_{h \to 0} \frac{f(A + hB) - f(A)}{h} B^{-1}$$

where $h$ is scalar.

In the following series of theorems, we shall, in each case assume the existence of $df(A)$ and $dA$.

**Theorem:**

$$\frac{d}{dA} [f(A) + g(A)] = \frac{df(A)}{dA} + \frac{dg(A)}{dA}$$

**Proof:**

$$\frac{d}{dA} [f(A) + g(A)]$$

$$= \lim_{h \to 0} \frac{f(A + hB) + g(A + hB) - f(A) - g(A)}{h} B^{-1}$$
\[
\lim_{h \to 0} \left[ \frac{f(A + hB) - f(A)}{h} \right] B^{-1} + \left[ \frac{g(A + hB) - g(A)}{h} \right] B^{-1}
\]

**Theorem:**

\[
\frac{d}{dA} \left[ \frac{f(A)g(A)}{dA} \right] = f(A)\frac{dg(A)}{dA} + g(A)\frac{df(A)}{dA}
\]

**Proof:** Let \( q(A) = f(A)g(A) \). Then:

\[
\frac{dq(A)}{dA} = \lim_{h \to 0} \left[ \frac{q(A + hB) - q(A)}{h} \right] B^{-1}
\]

\[
= \lim_{h \to 0} \left[ \frac{f(A + hB)g(A + hB) - f(A)g(A)}{h} \right] B^{-1}
\]

\[
= \lim_{h \to 0} \left[ \frac{f(A + hB)g(A + hB) - f(A + hB)g(A) + f(A + hB)g(A) - f(A)g(A)}{h} \right] B^{-1}
\]

\[
= \lim_{h \to 0} \left[ \frac{f(A + hB)[g(A + hB) - g(A)]}{h} \right] B^{-1}
\]

\[
+ \lim_{h \to 0} \left[ \frac{f(A + hB) - f(A)}{h} \right] g(A) B^{-1}
\]

and

\[
\frac{dq(A)}{dA} = f(A)\frac{dg(A)}{dA} + g(A)\frac{df(A)}{dA}
\]

since \( \frac{df(A)}{dA} \) does not exist if \( \lim_{h \to 0} \frac{f(A + hB) - f(A)}{h} \neq f(A) \).

**Theorem:**

\[
\frac{dA^m}{dA} = mA^{m-1}
\]

**Proof:**

\[
\frac{dA^m}{dA} = \lim_{h \to 0} \left[ \frac{(A + hB)^m - A^m}{h} \right] B^{-1}
\]

\[
= \lim_{h \to 0} \left[ \frac{A^m + mhA^{m-1}B + \cdots + m!B^m}{h} \right] B^{-1}
\]

\[
= \lim_{h \to 0} \left[ \frac{m^m - 1}{h} + h(m-1)(A^{m-2}B + \cdots + B^{m-2}) \right] B^{-1}
\]

\[
= mA^{m-1}
\]
It follows, in general, that
\[ \frac{d}{dA} (f(A))^n = n f(A) \left( f(A) \right)^{n-1} \]

where \( f(A) \) is a function of the matrix \( A \).

Note that the results in the preceding theorems are independent of \( B \) except for the condition, \( AB = BA \), where \( B \) is non-singular. Using the theorem on differentiation of sums, it follows that \( \frac{df(A)}{dA} \) is independent of \( B \) for any polynomial. In the future, then, we shall, without loss of generality, restrict ourselves to \( B = I \) when dealing with a polynomial. Furthermore, with this as motivation, we shall now adopt the following definition:

\[ \frac{df(A)}{dA} = \lim_{h \to 0} \frac{f(A + hI) - f(A)}{h} \]
CHAPTER V

POWER SERIES

The power series to be considered in this chapter are those for exponential functions, sines, and cosines.

**Definition:** Let $A$ be a matrix of order $n$. Then, the exponential function of $A$ is defined to be:

$$e^A = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots + \frac{A^n}{n!} + \cdots$$

We note that this infinite series corresponds to that for a scalar function. If $A$ is the zero matrix, $e^A = I = e(0)$.

**Definition:**

$$e^{-A} = I - A + \frac{A^2}{2!} - \frac{A^3}{3!} + \cdots + (-1)^n \frac{A^n}{n!} + \cdots$$

**Theorem:** The exponential series converges for all $A$.

**Proof:** An application of the theorem on a sufficient condition for convergence, it is seen that a dominant series for each of the series which are the elements of $e^A$ is the series expansion of:

$$\frac{n-1}{n} + \frac{c^n}{n}$$

where $U$ is the greatest modulus in $A$. We know that $e^k$ is convergent for all $k$ where $k$ is a scalar. Thus, $e^A$ converges for all $A$.

**Theorem:** Let $u$ be a matrix of order $n$ with distinct
characteristic roots $\lambda_i$, $i = 1, 2, \ldots, n$, and let $S(u)$ be an infinite matrix series such that $S(\lambda_1)$ is absolutely convergent. Then, if $t(u)$ is any rearrangement of $S(u)$,

$$\lim_{m \to \infty} t_m(u) = \lim_{m \to \infty} S_m(u) = S(u).$$

**Proof:** Let $t(\lambda_i)$ be any rearrangement of $S(\lambda_i)$. Let $t_m(\lambda_i)$ be the sum of the first $m$ terms of this rearrangement. It is well known for scalar series that:

$$\lim_{m \to \infty} t_m(\lambda_i) = S(\lambda_i).$$

Thus, for matrix series:

$$\lim_{m \to \infty} t_m(u) = \sum_{i=1}^{n} \lim_{m \to \infty} t_m(\lambda_i)z_i$$

$$= \sum_{i=1}^{n} S(\lambda_i)z_i = S(u),$$

where

$$z_i = \frac{\Pi_{j \neq i}(\lambda_j I - u)}{\Pi_{j \neq 1}(\lambda_j - \lambda_1)}$$

Let $R(u) = \sum_{i=0}^{\infty} a_i u^i$ and $S(u) = \sum_{i=0}^{\infty} b_i v^i$ where $u$ and $v$ are matrices of order $n$ with distinct characteristic roots, $\lambda_1$ and $\lambda_1$, respectively, $i = 1, 2, \ldots, n$. Let $u$ and $v$ be such that $uv = vu$ and $R(\lambda_1)$ and $S(\lambda_1)$ are absolutely convergent. Consider:

$$S(u, v) = a_0b_0 + a_1b_0u + a_1b_1uv + a_0b_1v + a_2b_0u^2 + a_2b_1uv + \ldots$$

so that:

$$S_m(u, v) = R_m(u)S_m(v).$$

It follows from the absolute convergence of $R(\lambda_1)$ and $S(\lambda_1)$
that the above series with \( u \) replaced by \( \lambda_i \) and \( v \) replaced by \( \frac{s_i}{i!} \) converges absolutely. Hence, we can rearrange the terms of \( S(u,v) \).

**Theorem:**
\[ e^{A+B} = e^A e^B , \] provided \( AB = BA \).

**Proof:** By multiplication of \( e^A \) and \( e^B \) according to the scheme used in the preceding argument, rearranging terms,
\[ (I + A + A^2 + \cdots)(I + B + B^2 + \cdots) \]
\[ = I + (A + B) + \left( \frac{A^2 + 2AB + B^2}{2!} \right) + \left( \frac{A^3 + 3A^2B + 3AB^2 + B^3}{3!} \right) + \cdots \]
but, \( I + B \) is again an \( n \times n \) matrix and by definition:
\[ e^{A+B} = I + (A + B) + \left( \frac{(A + B)^2}{2!} \right) + \left( \frac{(A + B)^3}{3!} \right) + \cdots \]

We note as a corollary, that, if \( u \) and \( v \) are polynomials of \( A \), \( e^{u+v} = e^v e^u = e^{u+v} \).

**Theorem:**
\[ e^A e^{-A} = e^{-A} e^A = I. \]

**Proof:** \( A \) and \(-A\) may be considered as polynomials in \( A \) and the theorem follows from the preceding one.

**Definition:** The matrix sine, \( \sin A \), where \( A \) is a matrix of order \( n \), is:
\[ \sin A \equiv A - \frac{1}{3!} A^3 + \frac{1}{5!} A^5 - \cdots \]

**Theorem:** \( \sin A \) converges for all \( A \).

**Proof:** Application of the theorem on a sufficient condition for convergence shows that a dominant series for each of the series which are elements of \( \sin A \) is:
where $U$ is the greatest modulus in $A$. We know that $\sin k$ is convergent for all $k$ where $k$ is a scalar. Thus, $\sin A$ converges for all $A$.

**Definition:** The matrix cosine, $\cos A$, is defined to be:

$$\cos A = I - \frac{1}{2!} A^2 + \frac{1}{4!} A^4 - ...$$

**Theorem:** $\cos A$ converges for all $A$.

**Proof:** By the theorem on a sufficient condition for convergence, a dominant series for every element of $\cos A$ is:

$$\frac{n - 1}{n} + \frac{1}{n} \left( I - \frac{1}{2!} (nu)^2 + \frac{1}{4!} (nu)^4 + ... \right)$$

$$= \frac{n - 1}{n} + \frac{1}{n} \cos(nU)$$

where $U$ is the greatest modulus in $A$. It is well known that $\cos k$ converges for all scalar $k$. Thus, $\cos A$ converges for all $A$.

**Theorem:**

$$\sin^2 A + \cos^2 A = I.$$

**Proof:** By multiplication of the series expansion of $\sin A$ according to the discussion and theorem in the first part of the chapter, and by rearranging the terms of the square:

$$\sin^2 A = A^2 - \frac{2A^4}{3^2} + \frac{16A^6}{3^5} - ...$$

By a similar argument:
Theorem: Let $t$ be a scalar variable and $u$ be a matrix of constants. Then,
\[
\frac{de_{ut}}{dt} = ue_{ut} = e_{ut}u
\]

Proof: From the proof of the convergence of $e^t$, it follows that $e^{ut}$ is absolutely and uniformly convergent for all $t$. The series obtained by differentiation of $e^{ut}$ term by term with respect to $t$ is
\[
u + u^2t + u^3t^2 + \ldots = ue^{ut}
\]
But, this series is absolutely and uniformly convergent. Thus, it represents the differential coefficient of $e^{ut}$. Hence,
\[
\frac{De_{ut}}{dt} = \frac{de_{ut}}{dt} = ue_{ut} = e_{ut}u
\]
where $D$ is the differential operator $\frac{d}{dt}$.

By repeated application of the preceding theorem,

\[
u^me_{ut} = u^me_{ut} = e_{ut}u^m.
\]

Theorem: Let $u$ be a matrix of order $n$ and let
\[
S = \lim_{n \to \infty} S_n
\]
where $S_n$ is a matrix series in the matrix $u$. Then, in any case for which we can extend Sylvester's Theorem, differentiation of $S$ term by term with respect to $u$ is justified.

Proof: By Sylvester's Theorem,
\[
S(u) = \sum_{r=1}^{n} S(\lambda_r) \sum_{j \neq r} \frac{1}{(\lambda_j I - u)_{j \neq r}} (\lambda_j I - u)
\]
\[ S(u+hI) = \sum_{r=1}^{n} S(\lambda_r + h) \prod_{j \neq r} \frac{(\lambda_j + h)(\lambda_j - \lambda_r)}{\lambda_j - \lambda_r} \]

since, if \( A \) is the matrix \( (u + hI) \), the characteristic roots of \( A \) are \( \lambda_r + h, r=1,2,\ldots,n \).

Then

\[ S(u+hI) - S(u) = \sum_{r=1}^{n} S(\lambda_r + h) \prod_{j \neq r} \frac{(\lambda_j + h)(\lambda_j - \lambda_r)}{\lambda_j - \lambda_r} \]

\[ - \sum_{r=1}^{n} S(\lambda_r) \prod_{j \neq r} \frac{\lambda_j I - u}{\lambda_j - \lambda_r} \]

\[ = \sum_{r=1}^{n} \left[ S(\lambda_r + h) - S(\lambda_r) \right] \prod_{j \neq r} \frac{\lambda_j I - u}{\lambda_j - \lambda_r} \]

Then:

\[ \lim_{h \to 0} \frac{S(u+hI) - S(u)}{h} = \]

\[ \sum_{r=1}^{n} \lim_{h \to 0} \frac{S(\lambda_r + h) - S(\lambda_r)}{h} \prod_{j \neq r} \frac{\lambda_j I - u}{\lambda_j - \lambda_r} \]

Let \( S'(\lambda_i) = \frac{dS(\lambda_i)}{d\lambda} \) at \( \lambda = \lambda_i \). Then,

\[ \frac{dS(u)}{du} = \sum_{r} \lim_{m \to \infty} \frac{d}{d\lambda} \left[ \prod_{j \neq r} \frac{\lambda_j I - u}{\lambda_j - \lambda_r} \right] \frac{\lambda_j I - u}{\lambda_j - \lambda_r} \]

\[ = \sum_{r} S'(\lambda_r) \prod_{j \neq r} \frac{\lambda_j I - u}{\lambda_j - \lambda_r} \]

If we let \( \varphi_r = \frac{\lambda_j I - u}{\lambda_j - \lambda_r} \), by Sylvester's Theorem:
Thus,

\[ S'(u) = \lim_{m \to \infty} S'_m(u). \]

**Theorem:** Let \( f(A) \) be a function of the matrix \( A \) such that the characteristic roots of \( f(A) \) are distinct, then,

\[ \frac{d e^{f(A)}}{dA} = \frac{df(A)}{dA} e^{f(A)}. \]

**Proof:** By definition:

\[ e^{f(A)} = I + f(A) + \frac{1}{2!} f(A)^2 + \frac{1}{3!} f(A)^3 + \ldots \]

\[ \frac{d e^{f(A)}}{dA} = \frac{df(A)}{dA} + 2f(A) \frac{df(A)}{dA} + \frac{3}{3!} f(A)^2 \frac{df(A)}{dA} + \ldots \]

\[ = \frac{df(A)}{dA} \left[ I + f(A) + \frac{1}{2!} f(A)^2 + \frac{1}{3!} f(A)^3 + \ldots \right] \]

**Theorem:** Let \( f(A) \) be a function of the matrix \( A \) such that the characteristic roots of \( f(A) \) are distinct, then,

\[ \frac{d \sin f(A)}{dA} = \frac{df(A)}{dA} \cos f(A). \]

**Proof:**

\[ \sin f(A) = f(A) - \frac{1}{3!} f(A)^3 + \frac{1}{5!} f(A)^5 - \ldots \]

\[ + \frac{(-1)^{n-1}}{(2n-1)!} f(A)^{2n-1} + \ldots, \quad n = 1, 2, \ldots \]
\[
\frac{dsin f(A)}{dA} = \frac{df(A)}{dA} - \frac{1}{3!} \frac{d^2 f(A)}{dA^2} + \frac{1}{5!} \frac{d^3 f(A)}{dA^3} - \ldots
\]

Theorem: Let \( f(A) \) be a function of the matrix \( A \) such that the characteristic roots of \( f(A) \) are distinct, then,

\[
\frac{dcos f(A)}{dA} = \frac{df(A)}{dA} \frac{dsin f(A)}{dA}
\]

Proof: Using the series expansion for \( \cos f(A) \) in place of that for \( \sin f(A) \) in the preceding proof, this theorem follows in a similar manner.
In this chapter, the symbol \( D \) will be used to denote a differential operator such as \( \frac{d}{dt} \). Thus,
\[
\frac{du}{dt} = Du \text{ and } \frac{d^2u}{dt^2} = D^2u.
\]

Consider the pair of first-order linear differential equations:
\[
\begin{align*}
\frac{v_1}{dt} y_1 + u_{11}y_1 + \frac{v_2}{dt} y_2 + u_{12}y_2 &= q_1 \\
\frac{v_2}{dt} y_1 + u_{21}y_1 + \frac{v_2}{dt} y_2 + u_{22}y_2 &= q_2
\end{align*}
\]
where \( v_{ij} \) and \( u_{ij} \) are constants and \( q_1 \) and \( q_2 \) are functions of \( t \). We may express these equations as:
\[
\begin{pmatrix}
\frac{v_1}{dt} + u_{11} & \frac{v_2}{dt} + u_{12} \\
\frac{v_2}{dt} + u_{21} & \frac{v_2}{dt} + u_{22}
\end{pmatrix}
\begin{pmatrix}
y_1 \\
y_2
\end{pmatrix}
=
\begin{pmatrix}
q_1 \\
q_2
\end{pmatrix}
\]
where \( D \) is the differential operator, \( \frac{d}{dt} \).

We call the matrix on the left an operational matrix. If we denote the operational matrix by \( f(D) \), we may write the set of differential equations as
\[
f(D) y = q
\]
where \( y = (y_1, y_2) \) and \( q = (q_1, q_2) \).

Definition: Let \( x_1, x_2, \ldots, x_n \) be \( n \) independent variables. Then, the matrix:
is called a matrix of partial differential operators.

We often find it necessary to differentiate bilinear or quadratic forms partially with respect to their variables.

Theorem: Let \( A(y,x) = yax = x^t a^t y \), where \( a \) is an \( n \times n \) matrix, \( y = (y_1, y_2, \ldots, y_n) \) and \( x = (x_1, x_2, \ldots, x_n) \).

Then, if:

\[
\left( \frac{\partial}{\partial y_j} \right) A(y,x) = (ax) \text{ and } \left( \frac{\partial}{\partial y_1} \right) A(y,x) = x^t a.
\]

Proof: We must keep in mind the fact that the quadratic and bilinear forms are scalar. Then, differentiate \( A(y,x) \) with respect to one of the variables \( y \), say, \( y = y_1 \).

Then:

\[
\frac{\partial A(y,x)}{\partial y_1} = (1, 0, 0, \ldots, 0)ax
\]

\[
= a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n. \]

If \( y = y_k \), it follows that:

\[
\frac{\partial A(y,x)}{\partial y_k} = (0, 0, \ldots, 1, 0, \ldots, 0)ax
\]

\[
= a_{11}x_1 + a_{21}x_2 + \ldots + a_{kn}x_n. \]

It follows from the preceding theorem that, if \( A(x,x) \) is the quadratic form \( x^t ax \), in which we assume \( a \) to be symmetric,

\[
\left( \frac{\partial}{\partial x_1} \right)^t A(x,x) = 2ax \text{ and } \left( \frac{\partial}{\partial x_1} \right) A(x,x) = 2x^t a,
\]

where

\[
\left( \frac{\partial}{\partial x_1} \right) = \begin{pmatrix} \frac{\partial}{\partial x_1} & \frac{\partial}{\partial x_2} & \cdots & \frac{\partial}{\partial x_n} \end{pmatrix}.
\]
Consider the formula, from the differential calculus of scalars, for the transformation from one set of \( n \) independent variables to another. If:
\[
\left( \frac{\partial}{\partial x_1} \right) = \left( \frac{\partial y_1}{\partial x_1}, \frac{\partial y_2}{\partial x_1}, \cdots, \frac{\partial y_n}{\partial x_1} \right)
\]
and
\[
\left( \frac{\partial}{\partial y_1} \right) = \left( \frac{\partial y_1}{\partial y_1}, \frac{\partial y_2}{\partial y_1}, \cdots, \frac{\partial y_n}{\partial y_1} \right)
\]
these relations may be expressed as:
\[
\left( \frac{\partial}{\partial x_1} \right)' = a' \left( \frac{\partial}{\partial y_1} \right)'
\]
where \( a \) is the matrix:
\[
a = \begin{pmatrix}
\frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\
\frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_n} \\
\cdots & \cdots & \cdots & \cdots \\
\frac{\partial y_n}{\partial x_1} & \frac{\partial y_n}{\partial x_2} & \cdots & \frac{\partial y_n}{\partial x_n}
\end{pmatrix}
\]
Similarly,
\[
\left( \frac{\partial}{\partial y_1} \right)' = b' \left( \frac{\partial}{\partial x_1} \right)'
\]
where \( b \) is the matrix obtained by interchanging \( x \) and \( y \) in \( a \).

**Theorem:**
\[
|a| |b| = 1
\]

**Proof:**
\[
\left( \frac{\partial}{\partial x_1} \right)' = a' b' \left( \frac{\partial}{\partial x_1} \right)'
\]
so that \( a' b' = 1 \) and \( a b = 1 \).
Thus
\[
|a| |b| = |1| = 1.
\]
We note that $a$ is the Jacobian, $\frac{\partial (y_1, y_2, \ldots, y_n)}{\partial (x_1, x_2, \ldots, x_n)}$ with which we are familiar from the calculus of scalar variables. $b$ is, then, the Jacobian for the transformation of an expression in terms of $y$ to one in terms of $x$. We note, also, that the Jacobian for one transformation multiplied by the Jacobian of the inverse transformation is 1.
CHAPTER VII

INTEGRATION

In this chapter, we shall only briefly consider integration of matrices with respect to a variable t since integration is, in general, the inverse operation of taking the derivative.

Definition: Let \( u \) be the matrix and \( t \) be a real variable such that:

\[
u = \begin{pmatrix} u_{ij}(t) \end{pmatrix}
\]

where the functions \( u_{ij}(t) \) are integrable with respect to \( t \).

Then,

\[
\int_{t_0}^{t} u dt = \int_{t_0}^{t} (u_{ij}(t))dt \equiv \left( \int_{t_0}^{t} u_{ij}(t)dt \right).
\]

We denote this integrated matrix by \( Q_{t_0}u \).

Theorem: Let \( u(t) \equiv (u_{ij}(t)) \) be of order \( n \). Then,

\[
Q_{t_0}u \leq U(t - t_0)K
\]

where the inequality sign means that each element of the matrix on the left is less than or equal to the corresponding element of the matrix on the right when \( U \) is the greatest modulus of all the elements of \( u(t) \) and \( K \) is a square matrix of order \( n \) having 1 for all its elements.

Proof:

\[
|u_{ij}(t)| \leq U.
\]
Then by a property of integrals, since $t$ is real,

$$
\int_{t_0}^{t} u_{ij}(t) \, dt \leq \int_{t_0}^{t} |u_{ij}(t)| \, dt \leq U(t - t_0).
$$

It follows, then, that every element of the matrix $C_{t_0}$ is less than or equal to $U(t - t_0)$. 
The problem of finding maximum likelihood estimates for parameters often arises in mathematical statistics. These maximum likelihood estimates are the values of the parameters for which the function distributing the random variables is a maximum. There are often a large number of these parameters, and, since we must solve the equations formed by setting the partial derivatives with respect to each of the parameters in order to maximize the function, it is often useful to write the function as a matrix function and then to apply matrices of differential operators to the matrix function. It is also common to maximize a function by maximizing the logarithm of the function.

Example: Suppose we are given the function:

\[ f(x_1, x_2) = 2a_1^2 x_1 + 3a_2^2 x_2 \]

and we wish to find the partial derivatives with respect to the parameters, \( a_1, a_2 \).

Let \( a = (a_1, a_2) \) and \( \Lambda = \begin{pmatrix} 2x_1 & 0 \\ 3x_2 & 0 \end{pmatrix} \)

where

\[ \frac{\partial}{\partial a_1} = \begin{pmatrix} \frac{\partial}{\partial a_1} & \frac{\partial}{\partial a_2} \end{pmatrix} \]
Then:
\[ f(x_1, x_2) = aAa' \]
and
\[ \left( \frac{\partial}{\partial x_{ij}} \right) f(x_1, x_2) = 2Aa' = \begin{pmatrix} 4a1x_1 \\ 6a2x_2 \end{pmatrix} \]
since \( aAa' \) may be considered a \( 1 \times 1 \) matrix.

**Example:** Consider the following problem from mathematical statistics. It is a part of the regression problem.

Let \( x_1, x_2, \ldots, x_n \) be \( n \) normally distributed random variables and let the expectation of \( x_i \) be \( E(x_i) \),
\[ E(x_i) = \alpha_1z_{i1} + \alpha_2z_{i2} + \cdots + \alpha_kz_{ik}. \]
Let the common variance be \( \sigma^2 \). Thus:
\[ f(x_1, x_2, \ldots, x_n) = \left( \frac{1}{\sigma^2} \right)^n e^{-\frac{1}{2} \sum_{i=1}^{n} (x_i - \sum_{j=1}^{k} \alpha_jz_{ji})^2} \]

Let:
\[ \log f(x_1, x_2, \ldots, x_n) = L \]
\[ = C - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \sum_{j=1}^{k} \alpha_jz_{ji})^2 \]
where \( C \) is some constant. Now let:
\[ \ell = (x_1, x_2, \ldots, x_n), \]
\[ \lambda = (\alpha_1, \alpha_2, \ldots, \alpha_k), \]
\[ Z = \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ z_{k1} & z_{k2} & \cdots & z_{kn} \end{pmatrix}, \]
\[ \left( \frac{\partial}{\partial \alpha_i} \right) = \begin{pmatrix} \frac{\partial}{\partial \alpha_1} \\ \frac{\partial}{\partial \alpha_2} \\ \vdots \\ \frac{\partial}{\partial \alpha_k} \end{pmatrix} \]
and
\[ K = \frac{n}{i=1} \left( z_{1} - \frac{k}{j=1} \alpha_{j} z_{j1} \right)^{2}. \]

We wish to find the values of \( \alpha_{1} \) which, for fixed \( z_{1} \) and \( z_{1} \), will make \( K \) take on its maximum value. These values will be denoted by \( \alpha_{1}^{*}, i = 1, 2, ..., k \). In order to find \( \alpha_{1}^{*} \) we solve simultaneously the set of equations which result when the partial derivatives with respect to \( \alpha_{1} \) are set equal to zero. Thus:

\[ K = (\boldsymbol{\xi} - \alpha \mathbf{Z})(\boldsymbol{\xi} - \alpha \mathbf{Z})' = (\boldsymbol{\xi} - \alpha \mathbf{Z})(\boldsymbol{\xi} - \alpha \mathbf{Z}') \]

\[ = \xi \xi' - \xi \mathbf{Z}' - \alpha \mathbf{Z} \xi' + \alpha \mathbf{Z} \mathbf{Z}' = \xi \xi' - 2 \alpha \mathbf{Z} \xi' + \alpha \mathbf{Z} \mathbf{Z}' \]

since \( \xi \mathbf{Z}' \) is a 1x1 matrix and thus:

\[ \xi \mathbf{Z}' = (\xi \mathbf{Z}')' = \mathbf{Z} \xi'. \]

Hence:

\[ \left( \frac{\partial}{\partial \alpha_{1}} \right)^{K} = \left( \frac{\partial}{\partial \alpha_{1}} \right)^{\xi \xi' - \alpha \mathbf{Z} \xi'} + \left( \frac{\partial}{\partial \alpha_{1}} \right)^{\alpha \mathbf{Z} \mathbf{Z}'} \]

\[ = - 2 \mathbf{Z} \xi' + 2 \mathbf{Z} \mathbf{Z}' \]

and

\[ - \mathbf{Z} \xi' + \alpha \mathbf{Z} \mathbf{Z}' = (0, 0, ..., 0)' \]

where \( \mathbf{Z} = (\alpha_{1}, \alpha_{2}, ..., \alpha_{k}) \)

Thus:

\[ z \xi' = 2 \mathbf{Z} \mathbf{Z}' \]

and

\[ z = (\mathbf{Z} \mathbf{Z}')^{-1} z \xi' \]
BIBLIOGRAPHY


