Robert Melville

$$\begin{pmatrix} 4 & 1 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 2 \\ -2 & 8 \end{pmatrix} = \begin{pmatrix} 2 & 8 \\ 3 & -2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 4 \\ 4 & 2 \end{pmatrix}\begin{pmatrix} 0 & 2 \\ 1 & -1 \end{pmatrix} \quad \begin{matrix} 5 & -2 \\ 2 & 6 \end{matrix}$$

MATRIX ANALYSIS I

MATHEMATICAL INTRODUCTION

1952

H. V. McIntosh

Table of Contents for Part I

A matrix is a rectangular array of mn numbers, called elements, arranged in m rows and n columns, which are usually enclosed within a large parenthesis, as

$$\begin{bmatrix} -3 & 1 & 2 \\ 6 & 0 & 4 \\ 2 & -1 & 7 \end{bmatrix}$$

A matrix may be designated by a single capital letter, say A. The various elements of A may be indicated by specifying the row and column in which they sit. Thus the ij^{th} element, called $[A]_{ij}$, is located at the intersection of the i^{th} row and j^{th} column.

The number of rows and columns in a matrix is called the order of the matrix. The matrix above is a 3 x 3 matrix since it has three rows and three columns.

Two matrices are considered to be equal when they are of the same order and their corresponding elements are equal. Thus

$$A = B \quad \text{when} \quad [A]_{ij} = [B]_{ij} \qquad \begin{matrix} i = 1, ---, n \\ j = 1, ---, m \end{matrix}$$

Two matrices of the same order may be added. By definition, this is done by adding the corresponding elements.

$$[A + B]_{ij} = [A]_{ij} + [B]_{ij}$$

or, to choose a specific example

$$\begin{bmatrix} -1 & 0 \\ 2 & -3 \end{bmatrix} + \begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 3 \\ 2 & -1 \end{bmatrix}$$

A matrix may be added to itself several times to obtain its integer multiples:

(-n summonds

$$A + A --- + A \quad = n\,A$$

so that $n[A]_{ij} = [nA]_{ij}$, which prompts the following definition for scalar multiplication

$$[kA]_{ij} = k[A]_{ij} = [Ak]_{ij}$$

where k is a complex number. These numbers are often called scalars to avoid confusion with matrices, which do not obey all the same arithmetical laws.

Since it is defined as an operation upon corresponding elements, it follows by direct verification that scalar multiplication for matrices obeys the associative, distributive, and commutative laws, and that scalar multiplication behaves the same as the multiplication of scalars among themselves.

The matrix, each of whose elements is zero, leaves unchanged any matrix of the same order to which it is added. Such a matrix is called an identity for addition, and is analogous to the scalar zero, which also has this property. Accordingly it is called a zero matrix and is designated by the symbol

$$\mathbb{O} \qquad\qquad [\mathbb{O}]_{ij} = 0$$

From a given matrix another may be formed by exchanging its rows for their corresponding columns. This process is called transposition, and the resulting matrix is the transpose of the original. A superscrip bar is used to designate the transpose, so that the transpose of A would be $\bar{A}$.

$$\bar{A}_{ij} = A_{ji}$$

first column

$$\overline{\begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 0 & 2 \end{bmatrix}} = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 1 & 2 \end{bmatrix}$$

becomes first row

The transpose of an mxn matrix is a nxm matrix. Furthermore, two transpositions in succession give back the original matrix.

$$\bar{\bar{A}} = A$$

If two matrices are equal, their transposes are equal, since in either case corresponding elements must be equal. The transpose of a sum of two matrices is the sum of their transposes, since

$$\left[\overline{A + B}\right]_{ij} = [A + B]_{ji} \quad [A]_{ji} + [B]_{ji}$$
$$= [\bar{A}]_{ij}{}^{2} \quad [\bar{B}]_{ij} = [\bar{A} + \bar{B}]_{ij}$$

thus $\overline{A + B} = \bar{A} + \bar{B}$

If the elements of a matrix are complex numbers, the complex conjugate of the matrix is gotten by taking the complex conjugate of each element. If a superscript star designates the complex conjugate,

$$[A^*]_{ij} = [A]_{ij}{}^*$$

and also $(A + B)^* = A^* + B^*$

These two operations, transposition and conjugation, permit a classification of matrices by certain symmetry properties. If a matrix is square, or in other words has the same number of rows as columns, it may happen that $\bar{A} = A$. When this is true, $[A]_{ij} = [A]_{ji}$ and the matrix is called symmetric. The origin of this name lies in the fact that if one draws in the main diagonal of the matrix

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

the matrix then is symmetric by mirror reflection in this diagonal. The main diagonal is also defined for non-square matrices, and passes through the elements of the form a_{ii}. Transposition is then the same as mirror reflection in this diagonal.

Then A has the form

$$A = \begin{bmatrix} \alpha & \beta & \gamma & \cdots \\ \beta & \delta & \cdots & \\ \gamma & \cdots & & \cdots \\ \cdots & & & \end{bmatrix}$$

On the other hand, the elements across the diagonal might be negatives of each other, and $\quad [A]_{ij} = -[A]_{ji}$

or

$$A = \begin{bmatrix} 0 & \beta & \gamma & \cdots \\ -\beta & 0 & \cdots & \\ -\gamma & \cdots & & \\ \cdots & & & \end{bmatrix}$$

Such a matrix is called antisymmetric, and $A = -\bar{A}$. The elements on the main diagonal must be zero, since $\quad [A]_{ii} = -[A]_{ii}$

From a given arbitrary square matrix M one may form a symmetric as well as an antisymmetric matrix:

$$M^{(s)} = \frac{1}{2}(M + \bar{M})$$

which is symmetric, since $\quad \bar{M}^{(s)} = \frac{1}{2}\overline{(M + \bar{M})} = \frac{1}{2}(\bar{M} + M)$; and

$$M^{(a)} = \frac{1}{2}(M - \bar{M})$$

which can be seen to be antisymmetric by similar reasoning. Since $M = M^{(s)} + M^{(a)}$, $M^{(s)}$ is called the symmetric part of M, and $M^{(a)}$ the antisymmetric part.

Similarly, if one has $A = \bar{A}^*$, A is said to be hermitean, while if $A = -\bar{A}^*$, it is antihermitean. If A is hermitean, the elements on the main diagonal must be purely real, since $[A]_{ij} = [A]_{ji}^*$, and images across the main diagonal are complex conjugates. Now, an antihermitean matrix may be written as i times a hermitean matrix, since if $\overline{iB}^* = iB$, $-i\bar{B}^* = iB$, $B = -\bar{B}^*$. An arbitrary square matrix may be separated into a hermitean part and an antihermitean part:

$$M^{(h)} = \frac{1}{2}(M + \bar{M}^*)$$

$$M^{(ah)} = \quad (M - \bar{M}^*)$$

so that $M = M^{(h)} + M^{(ah)}$. This process is analogous to writing a complex number as the sum of a real part and i times an imaginary part.

The elements of a matrix need not be complex numbers, but may be matrices themselves, for example. In particular, a matrix may be partitioned into smaller blocks, and these blocks be regarded as matrices. Thus if

$$M = \left[\begin{array}{cc|ccc} 1 & 2 & 1 & 0 & 6 \\ 3 & 1 & 2 & 6 & 5 \\ \hline 1 & 0 & 1 & 1 & 1 \\ 2 & 1 & 2 & 0 & 0 \end{array}\right]; \qquad N = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \qquad \mathbb{L} = \begin{bmatrix} 7 & 0 & 6 \\ 2 & 6 & 5 \end{bmatrix}$$

$$\mathbb{J} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \qquad \mathbb{T} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 0 \end{bmatrix}$$

then

$$M =$$

These small blocks are then called submatrices, while M is then a supermatrix. One case where this partitioning is useful lies in regarding a matrix as a row of columns, or alternatively, as a column of rows.

In transposing a supermatrix it does not suffice to merely exchange the rows for the columns; the elements themselves must be transposed, in addition. This can be most clearly seen by considering a matrix as a column of rows. If the elements of the supermatrix were not also transposed, the transpose of the supermatrix would be a row of rows, in other words a very long row, and not a row of columns as it should be.

Matrices having but one row or one column are called vectors. To distinguish the two cases they are referred to as row vectors and column vectors respectively. The transpose of a row vector is a column vector, and the transpose of a column vector is a row vector. For this reason it is customary to regard a vector as a column unless stated otherwise, and when it is necessary to discuss a row vector to write it as the transpose of the corresponding column. Vectors are usually designated by capital letters from the end of the alphabet. For example,

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ --- \\ x_n \end{bmatrix} \qquad \bar{X} = [x_1, x_2, x_3, ---, x_n]$$

Another notation is convenient for a column; it is called a "ket".

$$|x_i\rangle = \begin{bmatrix} x_1 \\ x_2 \\ --- \\ x_n \end{bmatrix}$$

The corresponding row is called a "bra", and is obtained from the ket by taking both the transpose and the complex conjugate

$$\langle x_i| = [x_1^*, x_2^*, --- x_n^*] = \overline{|x_i\rangle}^* \qquad \overline{|x_i\rangle}^*$$

Since the symbol is distinctive, this fact is indicated in no other way.

Since they are of sufficiently common occurrence, certain vectors are given a special symbol. These are the so-called coordinate vectors, which have a 1 in the i^{th} position and zeroes elsewhere. They are designated by $:i\rangle$ or $\langle i:$ accordingly as they are columns or rows. Thus

$$\langle 2: = [0 \ 1 \ 0 \ --- \ 0]$$

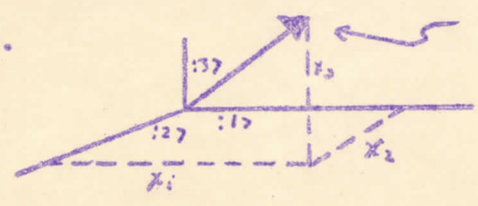
the reason that they are called coordinate vectors is that any vector of the proper order can be written as

$$|x_i\rangle = \sum_{(i)} x_i \ :i\rangle$$

or

$$\langle x_i| = \sum_{(i)} x_i^* \langle i:$$

If a vector is interpreted as providing the coordinates of a point in an n-dimentional space, these coordinate vectors point along the various axes of a cartesian coordinate system, and separate the various coordinates of the point.



$x_1 \langle 1\rangle + x_2 = :i\rangle + x_3 :3\rangle$

Because of this interpretation one sometimes speaks of the dimension of a vector rather than of its order as a matrix.

$$[AB]_{ij} = \sum_{K} [A]_{iK} [B]_{Kj}$$

$K=1$ Ai $A_{12}+B_{22}$

MATRIX MULTIPLICATION

$A_{11} + B_{11}$

Two matrices may be multiplied together according to the following rule:

$$[AB]_{ij} = \sum_{k} [A]_{ik} [B]_{kj}$$

In order to form such a product the number of columns of A must be the same as the number of rows of B. If A is of the order mxn and B is the order nxr, the product is of the order mxr. If two matrices have such an order that they may be multiplied together, they are called conformable. All the square matrices of a given order are conformable. $\begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$ 2

A row vector and a column vector of the same dimensionality are conformable, in which case the rule for a matrix product reduces to

ax y

$$\bar{X} Y = \sum_{k} x_k y_k \qquad \begin{pmatrix} a \\ b \end{pmatrix} (x \; y)$$

$+cy$

or

$$\bar{X} Y = [x_1, x_2, \cdots x_n] \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

A_{11} ax y

$AB_{11} =$

this product between two vectors is a scalar, and the product is called an inner product, to distinguish it from an outer product $X\bar{Y}$ which would be a matrix. The inner product is in a certain sense commutative, since

$$\bar{X} Y = \sum_{k} x_k y_k = \sum_{k} y_k x_k = \bar{Y} X$$

In each case the first vector of the product is transposed.

If one extends the concept of a three dimensional space to an n-dimensional space, he may generalize the pythagorean theorem to read

$$s^2 = x_1^2 + x_2^2 + \cdots + x_n^2$$

If a vector is regarded as a line segment joining the origin to the point $(x_1, x_2, \cdots, x_n)$, this formula gives the length of the diagonal of a rectangular hyperparallelopiped in terms of its sides, or alternatively, the

$(4\,1\,2)\begin{pmatrix}3\\6\\2\end{pmatrix} = 12+6+4=22$

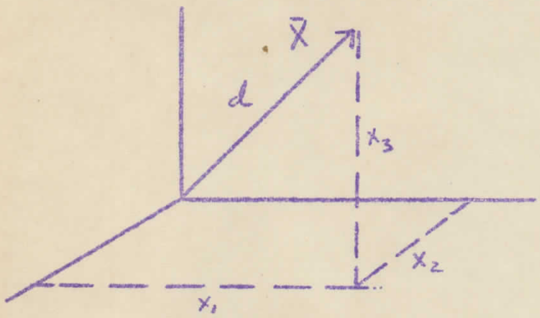
length of a vector in terms of its components.

This length, squared, may be written as

$$d^2 = x_1 x_2 + x_2 x_2 + \cdots + x_n x_n \quad (1\,2\,3)\begin{pmatrix}1\\2\\3\end{pmatrix}$$

so that if $[X]$ is the length of the vector X,

$$[X] = \sqrt{\bar{X}\,X}$$



$$d^2 = x_1^2 + x_2^2 + x_3^2 = \bar{X}\,X$$

Although such a definition of the length of a vector is useful, in terms of this geometrical picture, it may have certain embarrassing properties when the elements of the vector are complex numbers, so that, for instance

$$[\,1, i\,]\begin{bmatrix}1\\ i\end{bmatrix} = 1 - 1 = 0$$

and the "length" of a vector may very well turn out to be zero in spite of the fact that the vector itself is non-zero. Such a vector is called a null vector, in contrast to a zero vector, all of whose components are zero. Such a product, in fact, is not even guaranteed to be real, so that another definition is necessary for a length which will be positive, and zero only when the vector itself is zero. If z is a complex number, $z^* z$ has this property, so that the product

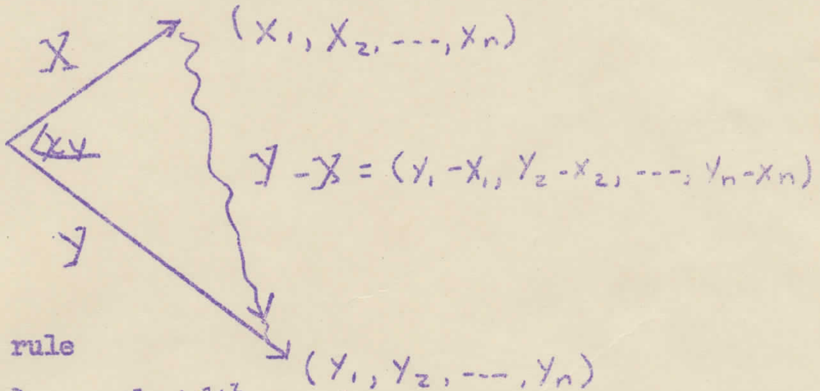
$$\langle x_i | x_i \rangle = [x_1^*, x_2^*, \ldots, x_n^*]\begin{bmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{bmatrix} = x_1^* x_1 + x_2^* x_2 + \cdots + x_n^* x_n$$

being the sum of positive or zero terms is itself positive or zero, and $\langle x_i | x_i \rangle \geq 0$. Furthermore it can only be zero when each term of the sum is zero, so that in that case $|x\rangle = 0$, a zero vector. This product is called the hermitean inner product, and furnishes the reason for which a bra was defined as the conjugate transpose of the corresponding ket.

The hermitean product of two vectors is not commutative, and in fact

$$\langle x|y\rangle = \sum x_i^* y_i = (\sum x_i y_i^*)^* = \langle y|x\rangle^*$$

To pursue the geometric analogy further, write the cosine law for n-dimensional space, taking for convenience the plane containing $(x_1, y_2 \cdots x_n)$, $(y_1, y_2, \cdots, y_n)$, and the origin.



by the cosine rule

$$|Y-X|^2 = |X|^2 + |Y|^2 - 2|Y||X| \cos \angle X,Y$$

$$\cos \angle X,Y = \frac{|x|^2 + |y|^2 - |y-x|^2}{2|x||y|}$$

$$= \frac{\bar{x}x + \bar{y}y - (\overline{y-x})(y-x)}{2|x||y|}$$

Now,

$$(\overline{y-x})(y-y) = (\bar{y}-\bar{x})(x-x) = \bar{y}y + \bar{x}x - \bar{y}x - \bar{x}y$$
$$= \bar{x}x + \bar{y}y - 2\bar{x}y$$

so that

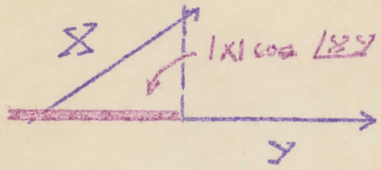
$$\cos \angle X,Y = \frac{\bar{x}y}{|x||y|}$$

which may be taken as the definity relation for this quantity, insofar as the geometry has been used only as an analogy.

This may be rewritten as

$$\bar{x}y = |X| \cdot |Y| \cos \angle X,Y$$

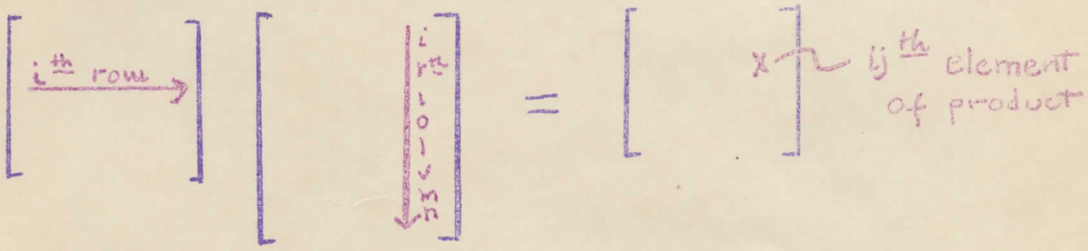
This provides a geometric interpretation for an inner product, for the quantity $\{|X|\cos \angle X,Y\}$ is just the projection of the vector X on the vector Y, so that the inner product is just this projection multiplied by the length of the vector . Obviously the roles of X and Y may be interchanged in this interpretation.



For the multiplication of two matrices, the first may be regarded as a supermatrix of rows, the second a supermatrix of columns. When this is done, the rule for the multiplication of matrices reads

$$[AB]_{ij} = \sum_k [A]_{ik}[B]_{kj}$$
$$= \sum_k [i^{th} \text{ row of } A]_k \, [j^{th} \text{ column of } B]_k$$

which is just the inner product of the i^{th} row by the j^{th} column. This gives a mnemonic device for performing matrix multiplication.



This gives the "over and down" rule; one starts at the left side of the i^{th} row of the first factor, the top of the j^{th} column of the second factor, as in the diagram above, multiplies the corresponding elements together, moves on to the next pair, and so on until all the corresponding pairs have been multiplied. These pairs are then added and their sum placed as the ij^{th} element of the product matrix. As an example

$$\begin{bmatrix} 1 & 2 \\ 0 & 4 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ -2 & 0 \end{bmatrix} = \begin{bmatrix} 1\cdot 1 + (2)(-2) & 1\cdot 1 + 2\cdot 0 \\ 0\cdot 1 + 4(-2) & 0\cdot 1 + 4\cdot 0 \end{bmatrix} = \begin{bmatrix} -3 & 1 \\ -8 & 0 \end{bmatrix}$$

The laws of multiplication, where they hold, must be verified for matrix multiplication. The following statements are true.

1. Matrix multiplication is associative.

$$A(BC) = (AB)C$$

so that the factors of a product may be grouped as desired, as long as their order is not changed.

Proof: $[A(BC)]_{ij} = \sum_k [A]_{ik}[BC]_{kj}$

$$= \sum_k [A]_{ik} \left\{ \sum_l [B]_{kl}[C]_{lj} \right\}$$

$$= \sum_k \sum_l [A]_{ik} [B]_{kl} [C]_{lj} \qquad 1)$$

while $[(AB)C]_{ij} = \sum_l [AB]_{il} [C]_{lj}$

$$= \sum_l \left\{ \sum_k [A]_{ik}[B]_{kl} \right\} [C]_{lj}$$

$$= \sum_l \sum_k [A]_{ik} [B]_{kl} [C]_{lj} \qquad 2)$$

comparing 1) and 2), they are seen to be the same except that the order of the summations has been changed, in one case the k summation coming first; in the other that over l. But these two procedures give the same result, as is seen by writing all the summands in the following array:

(constant k, increasing l)

constant l increasing k

$$\begin{array}{lll} [A]_{i1}[B]_{11}[C]_{1j} & [A]_{i1}[B]_{12}[C]_{2j} & --- \\ [A]_{i2}[B]_{21}[C]_{1j} & [A]_{i2}[B]_{22}[C]_{2j} & --- \\ --- & & \end{array}$$

In the case that the k summation is done first, the columns of the above array are first added, then the sums of columns are added by the l summation, altogether summing all the elements of the array. If the l summation is performed first, the elements sitting in the same row are added together first, then these sums by the k summation, again including all the numbers in the array. In either case the same terms are included in the sum, but in different orders. q.e.d.

2. Matrix multiplication is distributive:

$A(B+C) \qquad AB + AC \qquad ((l+m)n = ln + mn)$

In other words, parenthesis may be multiplied out. Proof:

$$\begin{aligned}[A(B+C)]_{ij} &= \sum_k [A]_{ik}[B+C]_{kj} \\ &= \sum_k [A]_{ik}\{[B]_{kj} + [C]_{kj}\} \\ &= \sum_k [A]_{ik}[B]_{kj} + \sum_k [A]_{ik}[C]_{kj} \\ &= [AB]_{ij} + [AC]_{ij}\end{aligned}$$

while a similar proof holds for right distributivity. q. e. d.

3. Matrix multiplication is not always commutative;

$AB \neq BA$ in general.

A counter-example is

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

so that multiplying two matrices in the two possible orders may very well give different results, depending upon the order. However matrix multiplication is not always non-commutative,

since

$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$$

The failure of the commutative law provides the reason for considering both a left and a right distributive law, but in both cases the order of the factors must be preserved.

Matrix multiplication thus proceeds as for scalars except that the order of the factors must be always preserved, in the cases where they are noncommutative.

Two further rules are useful:

4. $\overline{AB} = \bar{B}\,\bar{A}$

Proof:
$$\begin{aligned}[\overline{AB}]_{ij} &= [AB]_{ji} \\ &= \sum_k [A]_{jk}[B]_{ki} \\ &= \sum_k [\bar{A}]_{kj}[\bar{B}]_{ik} \\ &= [\bar{B}\bar{A}]_{ij}\end{aligned}$$

5. $(AB)^* = A^*B^*$

Proof:
$$\begin{aligned}[(AB)^*]_{ij} &= \left(\sum_k [A]_{ik}[B]_{kj}\right)^* \\ &= \sum_k [A]^*_{ik}[B]^*_{kj} \\ &= \sum_k [A^*]_{ik}[B^*]_{kj} \\ &= [A^*B^*]_{ij}\end{aligned}$$

It should be noted that the preceding proofs take for granted that the elements of the matrices obey the rules for arithmetic, as is the case when the elements are scalars. The fourth rule however involved the interchange of the order in which the elements were multiplied, so will be false when multiplying supermatrices whose elements fail to commute.

Corresponding to the 1 of scalar multiplication there are unit matricess for matrix multiplication, which are designated by the symbol $\mathbb{I}$, and which have the property that

$$\mathbb{I}A = A \qquad\qquad A\mathbb{I} = A$$

for each matrix A. Suppose $[\mathbb{I}]_{ij} = \delta_{ij}$; then the equation $\mathbb{I}A = A$ reads, elementwise:

$$\sum_k \delta_{ik}[A]_{kj} = [A]_{ij}$$

which, since the $[A]_{ij}$'s may be entirely arbitrary, can only be satisfied by

$$\delta_{ij} = \begin{matrix} 0 & i \neq j \\ 1 & i = j \end{matrix}$$

this symbol is called Kroneker's delta.

The unit matrices are necessarily square, since they give back a matrix of the same order as the one which they multiply.

A similar argument to the one above shows that the solution of the equation

$$A \mathbb{I} = A$$

for all A yields the same unit matrices. If A_{mxn} specified that A in an mxn matrix,

$$\mathbb{I}_{m \times m} \; A_{m \times n} \; \mathbb{I}_{n \times n} = A_{m \times n}$$

so that if A is not square, left and right units are actually different, being of different order. However if A is square, the units are the same, and in fact, a unit matrix commutes with each square matrix of its own order.

$$A \mathbb{I} = \mathbb{I} A = A$$

The product of a matrix and a column is again a column, although not necessarily of the same length. If

$$Y = A X$$
$$y_i = \sum_k a_{ik} x_k$$

A is then an operator which changes the vector X into the vector Y.

The quantity

$$\frac{\partial y_i}{\partial x_j}$$

is clearly seen to be a_{ij}

$$\frac{\partial y_i}{\partial x_j} = a_{ij}$$

and the term a_{ij} gives just the partial dependence of the element y_i upon the element x_j of the old vector. Furthermore a matrix is a linear operator, by the distributive law, since

$$A(W + Z) = AW + AZ$$

which may be taken as the definition of a linear operator; namely, one which operates upon sums in the same way that it operates upon the summands.

The product of two matrices operating upon a vector has the following interpretation. Let

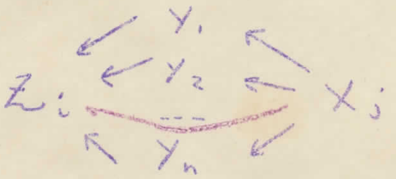
$$Y = AX$$
$$Z = BY$$

so that

$$Z = (BA)X$$

by the associative rule. Now the dependence of Z_i upon X_j has been obtained through an intermediate step, namely through considering the Y_k which depend upon X_j and then noting the dependence of Z_i upon the Y_k; altogether terms of the form

$$b_{ik} a_{kj}$$



to express the intermediary effect of the Y_k, and sum them to obtain the total dependence,

$$\sum_k b_{ik} a_{kj} = [BA]_{ij}$$

which coincides with the definition of a matrix product. Regarding the elements of the matrices as partial derivatives,

$$\sum_k b_{ik} a_{kj} = \sum_k \frac{\partial z_i}{\partial Y_k} \frac{\partial Y_k}{\partial x_j} = \frac{\partial z_i}{\partial x_j} = [BA]_{ij}$$

by the formula for the partial derivative of a function of a function.

Not only are matrices linear operators, but all linear operators on a vector may be written as matrices, at least when they are sufficiently continuous, for such operators may only involve the first powers of the elements of the vector, multiplied by linear operators and summed, so that these operators become just the elements of the matrix.

It was previously mentioned that the product $X\,Y$ between two vectors was called their outer product, in distinction to the product in the opposite order, which was called their inner product. Such a product is a matrix and has the property that each of the rows, regarded as a submatrix, is proportional to each of the others. The same is true of the columns.

$$X\bar{Y} = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} [y_1, y_2, \cdots y_n] = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_m \\ x_2 y_1 & x_2 y_{2}\ \text{2(?)} & \cdots & x_2 y_m \\ \cdots & & & \\ x_n y_1 & x_n y_2 & \cdots & x_n y_m \end{bmatrix}$$

When such a product is given, the factors may be recovered except for a multiplying factor. This is true since $(\frac{1}{a}X)(a\bar{Y})$ gives the same outer product as $X\bar{Y}$. Within this ambiguity the factors are recovered by noticing that the first factor is proportional to the columns of the outer product, and the second factor to the rows, and that if one selects the i^{th} row and the j^{th} column the product of these proportionality factors must be just $[X\bar{Y}]_{ij} = x_i y_j$

All these results concerning the multiplication of matrices may be summarized in the following table:

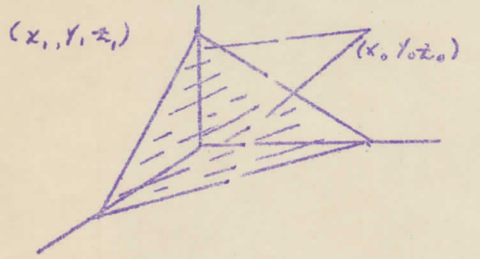
row x column = scalar	inner product
bra x ket = scalar	hermitean inner product
column x row = matrix	outer product
matrix x column = column	
row x matrix = row	
mxn matrices x nxr matrix = mxr matrix	

DETERMINANTS, INVERSES

The following examples have been chosen not only to illustrate some of the applications of vector methods in geometry, but because they lead to a cleraer insight into the nature of a matrix inverse.

Consider the problem of determining the distance from a point to a plane.

Let the point be (x_0, y_0, z_0), sitting in a three dimensional space, and let the equation of the plane be $Ax + By + Cz = D$. i_1, i_2, i_3 are unit vectors along the x-, y-, and z-axes respectively, and a vector



$X = (x_1 - x_0)i_1 + (y_1 - y_0)i_2 + (z_1 - z_0)i_3$ represents the line segment joining the point (x_1, y_1, z_1) of the plane to the given point.

Now, if there can be found a vector N', of length 1, perpendicular to the plane, then

$$\bar{X}N' = |X||N'| \cos \angle X, N'$$
$$= |X| \cos \angle X, N'$$

will give the perpendicular distance from the point to the plane. This normal vector, N' may be found by a consideration of the points (x_1, y_1, z_1) and (x_2, y_2, z_2), which lie in the plane.

$$Ax_1 + By_1 + Cz_1 = D$$
$$Ax_2 + By_2 + Cz_2 = D$$

where not all the coefficients A, B, C are zero. Subtracting these two equations:

$$A(y_1 - x_2) + B(y_1 - y_2) + C(z_1 - z_2) = 0$$

this is the inner product of the two vectors

$$\bar{N} = [A, B, C]$$

and
$$\bar{P} = [(x_1 - x_2), (y_1 - y_2), (z_1 - z_2)]$$

By hypothesis, $|N| \neq 0$, $|P| \neq 0$, so that the above equation

$$\bar{P}N = 0$$

means that $\cos \angle P, N = 0$ and $P \perp N$; the two vectors are perpendicular. However, P represents a line segment which points from any point of the plane to any other point, so that the only way that N can be perpendicular to all these vectors to be perpendicular to the plane itself. If N is divided by its own length, it becomes a vector of length 1.

$$N' = \frac{[A, B, C]}{\pm\sqrt{A^2+B^2+C^2}}$$

$$\bar{N}X = \frac{A(x_1-x_0)+B(y_1-y_0)+C(z_1-z_0)}{\pm\sqrt{A^2+B^2+C^2}}$$

Since $Ax_1+By_1+Cz_1 = D$

$$d = \bar{N}'X = \frac{Ax_0+By_0+Cz_0+D}{\pm\sqrt{A^2+B^2+C^2}}$$

which is a formula that is familiar from analytic geometry. Its generalization to spaces of different dimensionality is apparent.

If it is not a plane, but some other surface whose normal is desired, the same methods apply. Treating again the three dimensional case, consider a surface which is specified by

$$f(x, y, z) = \text{constant}$$

then,

$$df = 0$$

$$df = \frac{\partial f}{\partial x}dx + \frac{\partial f}{\partial y}dy + \frac{\partial f}{\partial z}dz$$

which formula is obtained by applying a limiting process to the methods of the previous example. As before, this expression can be factored,

$$\left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right]\begin{bmatrix} d_x \\ d_y \\ d_z \end{bmatrix} = 0$$

so that the vector $\left[\frac{\partial f}{\partial x}\ \frac{\partial f}{\partial y}\ \frac{\partial f}{\partial z}\right]$ is perpendicular to the vector $[dx, dy, dz] = \overline{dX}$ which is a small vector tangent to the surface. Hence,

$$\left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right] = \bar{N}'$$

is a vector normal to the surface. When f is the equation of a plane, this reduces to the previous case.

$\overline{N}'$ can be written in the form

$$\left[\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right] f$$

where the quantity $\frac{\partial}{\partial x} ; 1\rangle + \frac{\partial}{\partial y} : 2\rangle + \frac{\partial}{\partial z} ; 3\rangle$ is a vector differential operator, called "grad."

The corresponding column is not a transpose of "grad", so that the distinction is made by the special symbols

$$\backslash\!\nabla = \left[\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right] \qquad \nabla\!\!/\!/ = \begin{bmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{bmatrix}$$

while $\overline{\nabla\!\!/\!/} = \left[\frac{6}{x6}, \frac{6}{y6}, \frac{6}{z6}\right]$ and differentiates from the right, contrary to the usual behavior of differentiators.

The expression

$$df = 0$$

is then written as

$$\backslash\!\nabla f \cdot d\mathcal{X} = 0$$

and

$$df = \backslash\!\nabla f \cdot d\mathcal{X}$$

which is somewhat reminiscent of the relation

$$df(x) = f'(x)\, dx$$

and the fact the gradient of a function, as $\backslash\!\nabla f$ is called, may be interpreted as the vector derivative of a function. It gives the rate at which a function changes upon looking at different points. It is a vector perpendicular to the surfaces $f = \text{constant}$ and gives the maximum rate of change of the function for any choice of the direction of the differential $d\mathcal{X}$, since

$$df = |\backslash\!\nabla f|\ |d\mathcal{X}| \cos \angle \nabla f, d\mathcal{X}$$

and $\cos \angle \nabla f\, d\mathcal{X} \leq 1$ the equality holding only when $\angle \nabla f, d\mathcal{X} = 0, 180^\circ$; i.e. $\backslash\!\nabla f$ is parallel to $d\mathcal{X}$.

In the present example the gradient was taken by varying the vector from the origin to the point in question, and hence had the element $\partial/\partial x_i$.

it might be written symbolically as

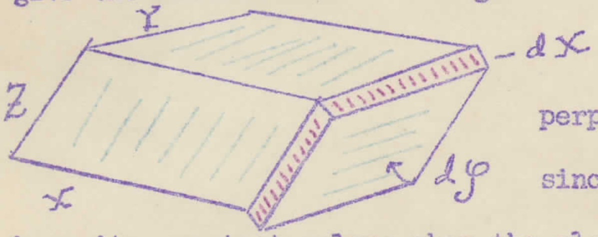
$$\nabla = \frac{d}{dX}$$

except that a) one cannot divide by a vector and b) the derivative depends upon the direction of dX

The vectors with respect to which a gradient may be taken are not limited to radial vectors, and a gradient might be formed with respect to any vector

$$\frac{d}{dY} = \left[\frac{\partial}{\partial y_1}, \frac{\partial}{\partial y_2}, ---, \frac{\partial}{\partial y_n}\right] = \nabla_Y$$

To illustrate this, consider a three dimensional parallelepiped which has adjacent edges the vectors X, Y, Z. The volume φ of this parallelepiped is then a function of these vectors, and $\nabla_X \varphi$, for instance, will give the maximum rate of change of the volume as the edge X is varied.



The gradient $\nabla_X \varphi$ will be a vector perpendicular to the face not containing X since a parallepiped with such a face will have its greatest volume when the slant height dX is perpendicular to the base, since the gradient points in the direction of the greatest rate of change of its function. The length of the gradient will be just the area of the base, since the change of the volume, $d\varphi$, is the area of the base times the differential slant height times the cosine of the angle between the vectors.

$$d\varphi = |dX| \, |\nabla_X \varphi| \cos \angle dX, \nabla_X \varphi$$

= (differential slant height) (base) (cosine of $\angle$)

= (base) (altitude of differential volume)

$\nabla_X \varphi$ is perpendicular to the base not containing X, and hence to Y and Z

$$\nabla_X \varphi \cdot Y = 0$$

$$\nabla_X \varphi \cdot Z = 0$$

and

$$\nabla_X \varphi \cdot X = \varphi$$

since the latter expression gives the base of the figure multiplied by the entire altitude, not just the differential altitude. The same is true for the other gradients, so that if it is observed that

$$\frac{1}{\varphi}\nabla_x \varphi = \nabla_x \ln \varphi$$

$$\begin{array}{lll} \nabla_x \ln\varphi \cdot X = 1 & \nabla_x \ln\varphi \cdot Y = 0 & \nabla_x \ln\varphi \cdot Z = 0 \\ \nabla_y \ln\varphi \cdot X = 0 & \nabla_y \ln\varphi \cdot Y = 1 & \nabla_y \ln\varphi \cdot Z = 0 \\ \nabla_z \ln\varphi \cdot X = 0 & \nabla_z \ln\varphi \cdot Y = 0 & \nabla_z \ln\varphi \cdot Z = 1 \end{array}$$

But this table of equations is the elementwise expression of the matrix equation:

$$\begin{bmatrix} \nabla_x \ln\varphi \\ \nabla_y \ln\varphi \\ \nabla_z \ln\varphi \end{bmatrix} [X, Y, Z] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Upon setting

$$M = [X, Y, Z] \qquad M^{-1} = \begin{bmatrix} \nabla_x \ln\varphi \\ \nabla_y \ln\varphi \\ \nabla_z \ln\varphi \end{bmatrix}$$

this equation reads

$$M^{-1}M = \mathrm{I}$$

A matrix M^{-1} having this property is called the inverse of M, and corresponds to the reciprocals among scalars. In particular such an inverse is called a left inverse because it sits on the left of the matrix M. But,

$$M^{-1}M = \mathrm{I}$$
$$\therefore \quad M(M^{-1}M) = M$$
$$(MM^{-1})M = M$$
$$\therefore \; MM^{-1} = \mathrm{I}$$

and the same matrix is also a right inverse. Thus

$$MM^{-1} = M^{-1}M = \mathrm{I}$$

and a square matrix commutes with its own inverse. Also,

$$(M^{-1})^{-1} M^{-1} = \mathbb{I}$$

so that $(M^{-1})^{-1} = M$, and the inverse of an inverse gives back the original matrix.

The inverse of a product of two matrices follows a special rule, namely $(AB)^{-1} = B^{-1}A^{-1}$. The proof is as follows:

$$(AB)(AB)^{-1} = \mathbb{I}$$
$$B(AB)^{-1} = A^{-1}$$
$$(AB)^{-1} = B^{-1}A^{-1}$$

where one has multiplied from the left successively by A^{-1} and B^{-1} and preserved this order on the right hand side of the equation.

While this argument for inverses has actually been carried out only for three dimensional matrices, it may be extended to the other cases. If M is a square matrix whose elements are $[M]_{ij}$

$$[M^{-1}]_{ij} = \frac{\partial \ln \mathcal{V}}{\partial [M]_{ji}}$$

where $\mathcal{V}$ is the volume of the hyperparallelepiped whose edges are the rows of M . Note that the derivative in this formula is taken with respect to $[M]_{ji}$ not $[M]_{ij}$. The reason for this is that the matrix multiplication is a row by column affair, while a row should be multiplied by the gradient with respect to a row, not a column. The transposition solves this difficulty.

If the quantity $\mathcal{V}$ vanishes the inverse does not exist and the matrix is then said to be singular.

$\mathcal{V}$, the volume of the n-dimensional parallelepiped whose edges are the rows of M , is called the determinant of M , and is symbolized by $|M|$.

It is written out in full as

$$\begin{vmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \cdots & & & \\ m_{n1} & m_{n2} & \cdots & m_{nn} \end{vmatrix}$$

to emphasize the fact that it is a function of the rows of M and in particular, of the elements of these rows.

Next, it is necessary to discover an explicit formula for evaluating the determinant. This may be done by making certain observations about volumes.

A determinant is a linear and homogeneous function of its rows. By homogeneous is meant that

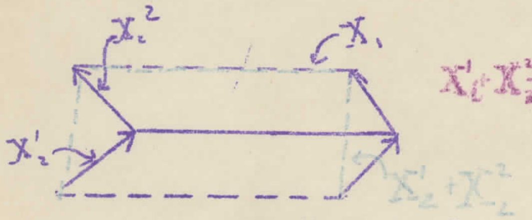
$$\varphi(\bar{x}_1, ---, k\bar{x}_i, ---\bar{x}_n) = k\varphi(\bar{x}_1, ---\bar{x}_n)$$

which is certainly true, since multiplying an edge of a parallelepiped by some factor multiplies the volume by that factor. By linear is meant that

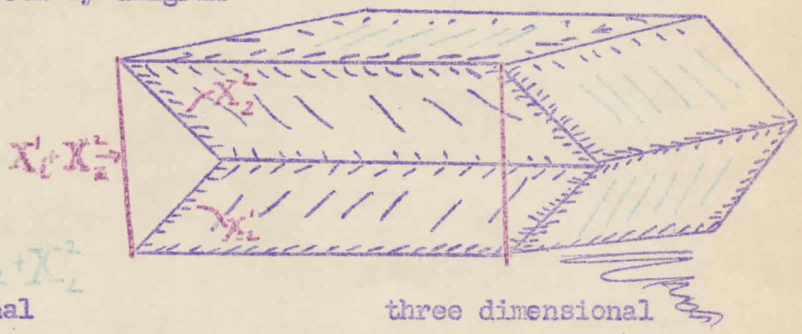
$$\varphi(\bar{x}_1, --- \bar{x}_i^1 + \bar{x}_i^2, ---, \bar{x}_n) =$$

$$\varphi(\bar{x}_1, ---, \bar{x}_i^1, ---, \bar{x}_n) + \varphi(\bar{x}_1, ---, \bar{x}_i^2 ---, \bar{x}_n)$$

the truth of which may be seen by diagram



two dimensional

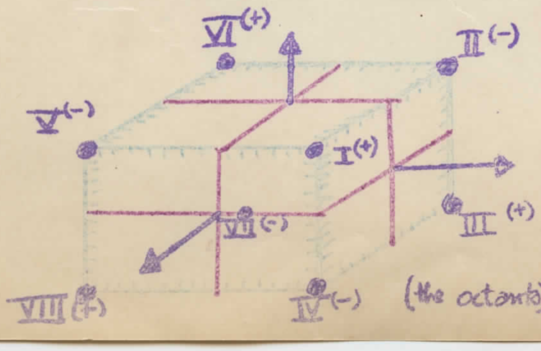


three dimensional

where the desired result may be seen by fitting the various prisms together to form the paralldepiped whose edges are the sums of the given edges.

The volume has a certain sign, since it is permitted that the edges be either positive or negative. In three dimensions for instance, if a certain cube sits in the first octant it has positive volume; negative in the second, etc. Also in an n-dimensional space there are 2^n-tants and a resulting handedness for the parallelepiped; changing 2^n-tants changes the sign; or changing sides changes the handedness and hence the sign; thus changing the order in which any two rows of the determinant appear changes its sign.

$$\varphi(---\bar{x}_i, ---\bar{x}_j, ---) = -\varphi(---\bar{x}_j, ---, \bar{x}_i---)$$



(the octants)

If two rows of the determinant are proportional, two adjacent sidesof the parallelepiped will be parallel and its volume will be zero. Furthermore, the cube which sits in the first 2^n-tant, edges along the coordinate axes, of unit length, has volume 1. So

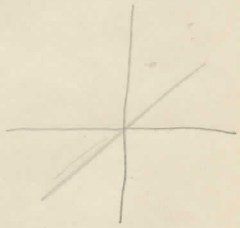
$$\varphi(\overline{\langle 1|}, \overline{\langle 2|}, ---, \overline{\langle n|}) = 1$$

with these four items of general information an explicit formula may be deduced.

Let $$|M| = \begin{vmatrix} m_{11} & m_{12} & --- & m_{1n} \\ m_{21} & m_{22} & --- & m_{2n} \\ --- & & & \\ m_{n1} & m_{n2} & --- & m_{nn} \end{vmatrix} = \begin{vmatrix} \sum m_{1i} \langle i| \\ \sum m_{2i} \langle i| \\ --- \\ \sum m_{ni} \langle i| \end{vmatrix}$$

$$= \sum_i \begin{vmatrix} m_{1i} \langle i| \\ \sum m_{2i} \langle j| \\ \sum --- \\ \sum_j m_{ni} \langle j| \end{vmatrix}$$ by linearity

$$\begin{vmatrix} 1 & 3 \\ 2 & 4 \end{vmatrix} = -2$$



$$= \sum_i \sum_j \begin{vmatrix} m_{1i} \langle i| \\ m_{2j} \langle j| \\ --- \\ \sum_l m_{nl} \langle l| \end{vmatrix}$$

$$= \sum_i \sum_j --- \sum_l \begin{vmatrix} m_{1i} \langle i| \\ m_{2j} \langle j| \\ --- \\ m_{nl} \langle l| \end{vmatrix}$$

$$= \sum_i \sum_j \cdots \sum_l m_{1i} m_{2j} \cdots m_{nl} \begin{vmatrix} \langle i| \\ \langle j| \\ --- \\ \langle l| \end{vmatrix}$$ by homogeneity

the determinant $\begin{vmatrix} \langle i| \\ \langle j| \\ --- \\ \langle l| \end{vmatrix}$ is of the form

$$\begin{vmatrix} 0 & --- & 1_{1i} & --- & 0 \\ 0 & -- & 1_{2j} & -- & 0 \\ -- & & & & \\ 0 & & -- & 1_{nl} -- & 0 \end{vmatrix}$$

and has zeroes or ones for its elements. This determinant will be zero unless each of the 1's occurs in a different column, otherwise the parallelepiped will be flat, and have zero volume.

If the 1's occur in different columns the rows can be rearranged to bring them onto the main diagonal. Each interchange of rows required to do this changes the sign of the volume once, and the volume is $+1$ when all the interchanges are accomplished. Thus

$$\begin{vmatrix} \angle i: \\ \angle j: \\ --- \\ \angle l: \end{vmatrix} = \begin{cases} (-1)^h & i \neq j, \neq k, \neq --, \neq l \quad el \\ 0 & \text{otherwise} \end{cases}$$

Altogether this gives

$$|M| = \sum_{(\text{permutations})} (-1)^h \, m_{1r_1} m_{2r_2} --- m_{nr_n}$$

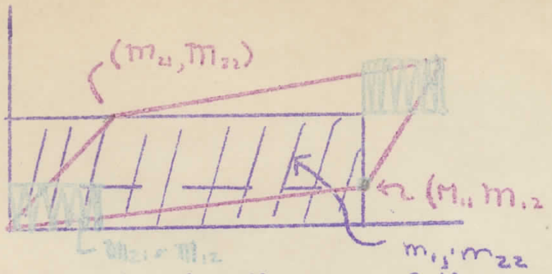
where $r_1 = 1, ---, n \qquad r_2 = 1, ---, n,$ except $r_1, \quad r_3 = ----$

In words, to evaluate the determinant $|M|$, select an element from the first row and some column; the second row and another column, etc. Multiply these elements together and prefix the sign $(-1)^h$ where h is the number of interchanges required to put the numbers $r_1, r_2, ---, r_n$ in the order 1, 2 ---, n, and add the results for each possible way to do this.

These are $n!$ summands, since there are n ways to select an element from the first row; $(n-1)$ from the second row and a different column, and so on, altogether $n(n-1) --- (n-n+1) = n!$

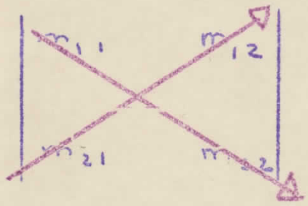
for a 1 x 1 matrix $\qquad |m| = m$

for a 2 x 2 $\qquad \begin{vmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{vmatrix} = m_{11} m_{12} -- m_{21} m_{12}$



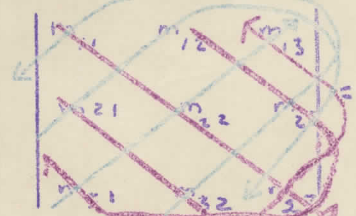
This formula receives its verification in the diagram at the left, where the areas may be matched up to give the area of the parallelogram.

There is a mnemonic device to evaluate a 2 x 2 determinant;

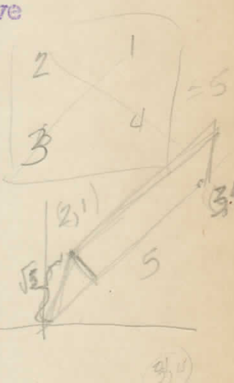


The elements on the downward slanting diagonal are multiplied together and given a $+$ sign; those on the upward slanting diagonal are multiplied and given a negative sign.

A 3 x 3 determinant may be evaluated by the same procedure.



$$= m_{11} m_{22} m_{33} + m_{12} m_{22} m_{31} + m_{13} m_{32} m_{21} - m_{31} m_{22} m_{13} - m_{32} m_{23} m_{11} - m_{21} m_{22} m_{33}$$



Save for the fact that there are many (24) parallelepipeds to juggle one could draw the same picture as for the 2 x 2 case.

This diagonal rule produces but $2n$ terms, so it fails when $2n \neq n!$, i.e. $n \neq 2, 3$. For a higher order determinant certain trickery must be resorted to, unless the rule derived above is to be applied directly. One important such rule is known as Laplace's development. In the formula

$$|M| = \sum_{\phi} (-1)^h m_{1r_1} m_{2r_2} \cdots m_{nr_n}$$

write

$$= \sum_{\phi} (m_{jr_j} (-1)^h m_{1r_1} \cdots m_{j-1, r_{j-1}} m_{j+1, r_{j+1}} \cdots m_{nr_n}$$

$$= \sum_{k} m_{jk} (-1)^{j+k} \sum_{\phi'} (-1)^{h'} m_{1r_1} \cdots m_{j-1\, r_{j-1}} m_{j+1, r_{j+1}} \cdots m_{nr_n}$$

$$= \sum_{k} m_{jk} \left\{ (-1)^{j+k} \mathcal{N}_{jk} \right\}$$

where the expression $\mu_{jk} = \sum_{p'} (-1)^{h'} m_{1r_1} \cdots m_{j-1\,r_{j-1}}\, m_{j+1\,r_{j+1}} \cdots m_{n r_n}$ is a sum with the terms m_{jr_j} deleted from the usual order. This is just the determinant which one would have if he crossed the j^{th} row and $r_j^{\,th}$ column from the original determinant:

$$\begin{vmatrix} m_{11} & m_{12} & \cdots & \cancel{m_{1r_j}} & \cdots & m_{1n} \\ m_{21} & m_{22} & & \cancel{m_{2r_j}} & \cdots & m_{2n} \\ \cancel{m_{j1}} & \cancel{m_{j2}} & \cancel{\cdots} & \cancel{m_{jr_j}} & \cancel{\cdots} & \cancel{m_{jn}} \\ m_{n1} & m_{n2} & & \cancel{m_{nr_j}} & \cdots & m_{nn} \end{vmatrix} = \mu_{j r_j}$$

since such elements have been removed from consideration. Such a determinant, μ_{jr_j} is called the "minor" of the element m_{jr_j}. To see that the signs of the above expressions are correct, notice that in calculating the minor the indices of the elements do not have the arrangement $1, 2, \cdots n$; but rather the first indices are ordered $1, 2, \cdots j-1, j+1, \cdots n$, while the second indices are ordered $r_1 \cdots r_{j-1}\ r_{j+1} \cdots n$ h' measures just the number of interchanges required to attain this order, while to attain the order $1, 2, \cdots n$, requireing th interchanges requires that the additional element m_{jr_j}, be inserted, say at the front of the sequence, and $(j+r_j)$ additional interchanges be made to place the two indices in the right order.

The quantity $(-1)^{j+k}\mu_{jk}$, the minor with a sign prefixed is called the cofactor of the element m_{jk}. Let

$$(-1)^{j+k}\mu_{jk} = \mathcal{M}_{jk}$$

then

$$|M| = \sum_k m_{jk}\,\mathcal{M}_{jk}$$

this expansion of the determinant in terms of the elements of some row is referred to as Laplace's development, and furnishes a systematic process for writing a determinant of a certain order in terms of determinants of lesser order. For example, use a 4 x 4 matrix

$$\begin{vmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \end{vmatrix} \qquad m_{11}\begin{vmatrix} m_{21} & m_{23} & m_{33} \\ m_{32} & m_{33} & m_{34} \\ m_{42} & m_{43} & m_{44} \end{vmatrix} - m_{12}\begin{vmatrix} m_{21} & m_{22} & m_{24} \\ m_{31} & m_{33} & m_{34} \\ m_{41} & m_{43} & m_{44} \end{vmatrix}$$

$$+ m_{13}\begin{vmatrix} m_{21} & m_{22} & m_{24} \\ m_{21} & m_{32} & m_{34} \\ m_{41} & m_{42} & m_{44} \end{vmatrix} - m_{14}\begin{vmatrix} m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \\ m_{41} & m_{42} & m_{43} \end{vmatrix}$$

which is the expansion of its determinant according to the first row.

Recalling the formula

$$|M| = \nabla_{x_j}|M| = x_j$$
$$= \Sigma \left[\nabla_{x_j}|M|\right]_k m_{jk}$$

and comparing with the formula of the last page

$$\frac{\partial |M|}{\partial m_{jk}} = \left[\nabla_{x_j}|M|\right]_k = M_{jk}$$

this formula may also be verified directly from Laplace's development, and shows that the partial derivative of a determinant with respect to an element is the cofactor of that element.

Geometrically this development amounts to writing the height of the parallelepiped in terms of the coordinate vectors, and using the fact that the volume is linear with respect to the edge vectors to write it as a sum of volumes of parallelepiped whose altitudes lie along the coordinate axes. Since the part of the bases parallel to the altitude does not contribute to the volume, the j^{th} coordinates of the vector of the base can be neglected when considering the altitude in the j^{th} direction, which corresponds to forming the minor by crossing out the j^{th} column, as well as the k^{th} row which contributes the altitudes. But this is again the volume of a base which is a volume in $(n-1)$ dimensional space, hence again a determinant. However the sign of the base, determined by its handedness with respect to the original parallelepiped is taken into account by the factor $(-1)^{j+k}$

The following checkerboard design is helpful for picking out the correct sign for the cofactor. The sign sitting where the element was chosen is prefixed to its minor to give the cofactor.

$$\begin{vmatrix} + & - & + & \cdots \\ - & + & - & \cdots \\ + & - & + & \cdots \\ \cdots & & & \end{vmatrix}$$

Needless to say, when but one element m of a given row of a determinant is non-zero, Laplace's method enables one to immediately reduce the determinant's order by 1 by writing $|M| = m$ (cofactor).

A number of properties of determinants, some of which have already been mentioned, follow immediately from either the formula, the use of Laplace's development, or the geometric picture. These properties are all of frequent commercial use.

1. If all the elements of a row are zero, the determinant is zero.

2. If two rows are the same, the determinant is zero.

3. Upon interchanging two rows of the determinant, the sign of the determinant changes. This furnishes a proof of property two. Changing the rows leaves the determinant unaltered, yet by 3) the sign changes $|M| = -|M|$ and $|M| = 0$.

4. If a row is multiplied by a factor c, the entire determinant is multiplied by this factor. This property was originally postulated for determinants to find their explicit formulae. This permits the removal of a constant factor from a row, which is useful numerically in keeping the elements of the determinant small. Note that this property differs from the corresponding property of matrices whereby to multiply a matrix by a scalar each element of the matrix is multiplied by that scalar. If each element of a determinant is multiplied by c, the determinant is multiplied by c^n.

5. A determinant may be written as the sum of two determinants which

are identical save for the elements of one row, the sum of which rows is the corresponding row of the original determinant. This is the linearity property originally postulated.

6. To any row of a determinant may be added a multiple of any other row without changing the value of the determinant. Writing the result as the sum of two determinants, one is zero since upon factoring out the multiplier, two of its rows will be identical, while the other determinant is the original determinant.

The use of these properties is illustrated in the evaluation of the following determinant:

3
+ row 4

$$\begin{vmatrix} 1 & 2 & 6 & 1 \\ 0 & -3 & 1 & 4 \\ -1 & 0 & -5 & 0 \\ 3 & 2 & 1 & 1 \end{vmatrix} = \begin{vmatrix} 0 & 2 & 1 & 1 \\ 0 & -3 & 1 & 4 \\ -1 & 0 & 5 & 0 \\ 3 & 2 & 1 & 1 \end{vmatrix}$$

− row 1

$$= \begin{vmatrix} 0 & 2 & 1 & 1 \\ 0 & -3 & 1 & 4 \\ -1 & 0 & 5 & 0 \\ 3 & 0 & 0 & 0 \end{vmatrix}$$

$$= (-1)^{1+4}\, 3 \begin{vmatrix} 2 & 1 & 1 \\ -3 & 1 & 4 \\ 0 & 5 & 0 \end{vmatrix}$$

$$= (-3)(-5) \begin{vmatrix} 2 & 1 \\ -3 & 4 \end{vmatrix}$$

$$= 15 \cdot 11$$

$$= 165$$

A systematic numerical method for evaluating a determinant is provided by the so-called reduction to triangular form: given the determinant

$$\begin{vmatrix} m_{11} & m_{12} & --- & m_{1n} \\ m_{21} & m_{22} & -- & m_{2n} \\ --- & & & \\ m_{n1} & m_{n2} & -- & m_{nn} \end{vmatrix}$$

Leave the first row unchanged; multiply the second row by $\dfrac{m_{11}}{m_{21}}$, the third by $\dfrac{m_{11}}{m_{31}}$, etc.

$$\begin{vmatrix} m_{11} & m_{12} --- & m_{1n} \\ m_{11} & \dfrac{m_{22}}{m_{21}} m_{11} -- & \dfrac{m_{2n}}{m_{21}} m_{21} \\ --- & & \\ m_{11} & \dfrac{m_{n2}}{m_{n1}} m_{11} --- & \dfrac{m_{nn}}{m_{n1}} m_{11} \end{vmatrix}$$

Subtract the first row from each of the others.

$$\begin{matrix} m_{11} & m_{12} & --- & m_{1n} \\ 0 & \dfrac{m_{22} m_{21} - m_{12} m_{21}}{m_{21}} & --- & \dfrac{m_{2n} m_{11} - m_{1n} m_{21}}{m_{21}} \\ --- & & & \\ 0 & \dfrac{m_{n2} m_{11} - m_{12} m_{n1}}{m_{n1}} & & \dfrac{m_{nn} m_{11} - m_{n1} m_{1n}}{m_{n1}} \end{matrix}$$

Repeat the process with the minor of m_{11} and so on until the determinant is brought to the form

$$\begin{vmatrix} m_{11} & m_{11} & -- & m_{1n} \\ 0 & \dfrac{m_{21} m_{11} - m_{12} m_{21}}{m_{21}} & & \dfrac{m_{21} m_{11} - m_{1n} m_{21}}{m_{21}} \\ 0 & 0 & \alpha --- & \beta \\ 0 & 0 & 0 --- & \gamma \end{vmatrix}$$

By Laplace's development, the value of the determinant is now just the product of the diagonal elements of the determinant.

Upon considering the formulae

$$|M| = \sum_{p} (-1)^h m_{1r_1} m_{2r_2} \cdots m_{nr_n}$$

$$|\bar{M}| = \sum_{p} (-1)^h m_{r_1 1} m_{r_2 2} \cdots m_{r_n n}$$

it is seen that they are the same: $|M| = |\bar{M}|$. In each case one selects an element from each row and column, and the number of interchanges needed to put the second indices in the order $1, 2, \cdots, n$ when the first indices are already in that order is just the same as to put the first indices in that order when the second indices have already been so arranged. Hence the signs of the individual temrs are the same by either scheme, and eventually the same terms are summed.

It may also be noted that

$$|M|^* = (\sum (-1)^h m_{1r_1} m_{2r_2} \cdots m_{nr_n})^*$$

$$= \sum (-1)^h m^*_{1r_1} m^*_{2r_2} \cdots m_{nr_n}{}^*)$$

$$= |M^*|$$

Geometrically speaking, one obtains the formula for $|M|$ by juggling the edges of the parallelepiped whose volume it expresses. However, the same arguments hold when one performs the same operations upon the coordinates, and these result in operations on the columns of the matrix, i.e. upon the rows of the transposed matrix.

Furthermore, each of the rules and deductions previously made hold for the columns as well as the rows, a conclusion which follows from consideration of the transposed matrix. In particular, Laplace's development holds by columns also;

$$|M| = \sum_{i} m_{ij} \mathfrak{M}_{ij}$$

a conclusion which also follows from the fact that $M^{-1}M = \mathrm{I\!I}$ as well as $M\,M^{-1} = \mathrm{I\!I}$

In one case one multiples the rows of M by the columns of M^{-1}, and hence by the cofactors of the row; in the other case he multiplies the rows of M^{-1} and thus the cofactor of the columns of M by the columns of M.

COORDINATE SYSTEM AND MORE DETERMINANTS

The formula for a matrix inverse may be applied to the solution of a system of simultaneous linear equations. Let there be equations in unknowns.

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = c_1$$

$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = c_2$$

$$\cdots$$

$$a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = c_n$$

where the indices of the constants a_{ij} tell in which equations and of what variable they appear as coefficients.

These equations may be written in the form

$$\sum a_{ij}x_j = c_i$$

and as such give the elements of the matrix product

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \cdots \\ c_n \end{bmatrix}$$

$$AX = Y$$

The equations are solved by expressing the x's in terms of the c's. If there is an inverse matrix A^{-1},

$$X = A^{-1}Y$$

by premultiplication of the previous equations by A^{-1}.

$$x_1 = \alpha_{11} c_1 + \alpha_{12} c_2 + \cdots + \alpha_{1n} c_n$$
$$x_2 = \alpha_{21} c_2 + \alpha_{22} c_2 + \cdots + \alpha_{2n} c_n$$
$$\cdots$$
$$x_n = \alpha_{n1} c_1 + \alpha_{n2} c_2 + \cdots + \alpha_{nn} c_n$$

where

$$\alpha_{ij} = \partial \ln|A|/\partial a_{ji} = \frac{1}{|A|}\,\partial|A|/\partial a_{ji} = \frac{1}{|A|}\,a_{ji}$$

with a_{ji} as the cofactor of a_{ji}. Then

$$x_i = \sum_j \frac{1}{|A|}\, a_{ji}\, c_j$$
$$= \frac{1}{|A|} \sum_j a_{ji} \cdot c_j$$

the term $\sum_j a_{ji} c_j$ is nothing other than another determinant, whose i^{th} column has the c's, not the a's as elements

$$X_i = \frac{\begin{vmatrix} a_{11} & a_{12} & \cdots & c_1 & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & c_2 & \cdots & a_{2n} \\ \cdots \\ a_{n1} & a_{n2} & \cdots & c_n & \cdots & a_{nn} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}}$$

This formula expresses Cramer's rule.

The condition that a system of equations have a solution is that the matrix of their coefficients have a non-vanishing determinant; or in other words that this matrix be non-singular.

An important question which can arise, when $Y = 0$, is whether there are any non-zero vectors X such that $AX = 0$. The answer will be no, if $|A| \neq 0$ for multiplying the equations by A^{-1}

$$X = A^{-1}\,0 = 0$$

On the other hand, there will be such a vector when $|A| = 0$, as may be seen by casting the product AX in the proper form: Considering the rows of A as submatrices,

$$AX = \begin{bmatrix} \text{row}1 \cdot X \\ \text{row}\,2 \cdot X \\ \cdots \\ \text{row}\,n \cdot X \end{bmatrix}$$

If the determinant is zero, it means that the volume spanned by the rows of A is zero, and hence that all these vectors lie in at worst an $(n-1)$ dimensional subspace. If the vector X is then chosen perpendicular to this space, its projection on all the rows will be zero, and AX will then be a zero vector.

Suppose that the elements of a determinant are functions of a parameter. Thus

$$|M| = |M(t)|$$

to differentiate the determinant with respect to the parameter, is to find

$$d\,|M|/dt$$

one may apply directly the firmula

$$|M| = \sum_p (-1)^h m_{1r_1} m_{2r_2} - \; m_{nr_n}$$

then

$$\frac{d|M|}{dt} = \sum_p (-1)^h \frac{dm_{1r_1}}{dt} m_{2r_2} \cdots m_{nr_n}$$
$$+ \sum_p (-1)^h m_{1r_1} \frac{dm_{2r_2}}{dt} \cdots m_{nr_n}$$
$$+ \cdots$$

giving a sum of determinants:

$$\frac{d}{dt}\begin{vmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \cdots & & & \\ m_{n1} & m_{n2} & \cdots & m_{nn} \end{vmatrix} = \begin{vmatrix} \frac{dm_{11}}{dt} & \frac{dm_{12}}{dt} & \cdots & \frac{dm_{1n}}{dt} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \cdots & & & \\ m_{n1} & m_{n1} & \cdots & m_{nn} \end{vmatrix}$$
$$+ \begin{vmatrix} m_{11} & m_{12} & \cdots & m_{2n} \\ \frac{dm_{21}}{dt} & \frac{dm_{22}}{dt} & \cdots & \frac{dm_{2n}}{dt} \\ \cdots & & & \\ m_{n1} & m_{n2} & \cdots & m_{nn} \end{vmatrix}$$
$$+ \cdots$$
$$+ \begin{vmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ \cdots & & & \\ \frac{dm_{n1}}{dt} & \frac{dm_{n2}}{dt} & \cdots & \frac{dm_{nn}}{dt} \end{vmatrix}$$

A similar formula, in which one column at a time had been differentiated would also hold, as well as more involved formulae.

Another, rather useful property of determinants, is that two determinants of the same order may be multiplied like matrices.

$$|BA| = |B|\ |A| = |AB|$$

And, since $|A| = |\bar{A}|$, three formulae hold in addition:

$$|A\bar{B}| = |A|\ |B|$$

$$|\bar{A}\bar{B}| = |A|\,|B|$$

$$|\bar{A}B| = |A|\ |B|$$

To prove the first assertion, consider the determinant of the supermatrix

$$\begin{bmatrix} A & -I \\ 0 & B \end{bmatrix} = N$$

In the first place, $(N) \cdot |A|\,|B|$, which can be seen by bringing the whole determinant to the triangular form.

$$|N| = \begin{vmatrix} A & -I \\ 0 & B \end{vmatrix}$$

$$= \begin{vmatrix} a_{11} & a_{12} & --- & a_{1n} & -1 & 0 & -- & 0 \\ a_{21} & a_{22} & -- & a_{2n} & 0 & -1 & -- & 0 \\ --- & & & -- & & -- & & \\ a_{n1} & a_{n2} & -- & a_{nn} & 0 & 0 & -- & -1 \\ 0 & 0 & -- & 0 & b_{11} & b_{12} & -- & b_{1n} \\ 0 & 0 & -- & 0 & b_{21} & b_{22} & -- & b_{2n} \\ - & - & & -- & & -- & & \\ 0 & 0 & -- & 0 & b_{n1} & b_{n2} & -- & b_{nn} \end{vmatrix}$$

The diagonal blocks A and B can obviously be brought to triangular form separately, due to the presence of 0 in the lower left hand corner, thereby bringing (N) to triangular form. The determinant is then just the product of the diagonal elements, and the product may be written in two parts, one of which will be $|A|$, the other $|B|$. Thus $|N| = |A||B|$

The geometrical interpretation of this maneuver is that two sets of edges, namely $[A, -\mathrm{I}]$ and $[0, B]$ have some mutually perpendicular components. Thus if $[0, B]$ is taken as a base for the parallelepiped, only the components of the base $[A, -\mathrm{I}]$ perpendicular to it will contribute to the volume, and this represents just the vectors $[A]$, so that the product of the "area" of the base $[A]$ with the base $[B]$ will give the volume of the entire figure.

But

$$|N| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} & -1 & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & a_{2n} & 0 & -1 & \cdots & 0 \\ \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & 0 & 0 & \cdots & -1 \\ 0 & 0 & \cdots & 0 & b_{11} & b_{12} & \cdots & b_{1n} \\ 0 & 0 & \cdots & 0 & b_{21} & b_{22} & \cdots & b_{2n} \\ \cdots \\ 0 & 0 & \cdots & 0 & b_{n1} & b_{n2} & \cdots & b_{nn} \end{vmatrix}$$

multiply the first row by b_{11}, the second by b_{12}, the n^{th} by b_{in}, and add all these rows to the $(n+1)$ st

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} & -1 & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & a_{2n} & 0 & -1 & \cdots & 0 \\ \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & 0 & 0 & \cdots & -1 \\ c_{11} & c_{12} & \cdots & c_{1n} & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & b_{21} & b_{22} & \cdots & b_{2n} \\ \cdots \\ 0 & 0 & \cdots & 0 & b_{n1} & b_{n2} & \cdots & b_{nn} \end{vmatrix}$$

where

$$c_{11} = a_{11} b_{11} + a_{12} b_{12} + \cdots + a_{n1} b_{1n}$$
$$c_{12} = a_{12} b_{12} + a_{22} b_{22} + \cdots + a_{n2} b_{1n}$$
$$\cdots$$
$$c_{1n} = a_{1n} b_{11} + a_{2n} b_{12} + \cdots + a_{n1} b_{1n}$$

or
$$c_{1j} = \sum_k a_{kj} b_{1k}$$
$$= [BA]_{1j}$$

repeat the process, now multiplying the first row by b_{21}, the second by b_{22}, ---, , the nth by b_{2n}

$$|N| = \begin{vmatrix} a_{11} & a_{12} & -- & a_{1n} & -1 & 0 & -- & 0 \\ a_{21} & a_{22} & -- & a_{2n} & 0 & -1 & -- & 0 \\ \text{---} & \text{---} & & & & & & \\ a_{n1} & a_{n2} & -- & a_{nn} & 0 & 0 & -- & -1 \\ c_{11} & c_{12} & -- & c_{1n} & 0 & 0 & -- & 0 \\ c_{21} & c_{22} & -- & c_{2n} & 0 & 0 & -- & 0 \\ 0 & 0 & -- & 0 & b_{31} & b_{32} & -- & b_{3n} \\ \text{---} & \text{---} & \text{---} & & & & & \\ 0 & 0 & -- & 0 & b_{n1} & b_{n2} & -- & b_{nn} \end{vmatrix}$$

$$c_{2j} = [BA]_{2j}$$

repeating until there remains

$$|N| = \begin{vmatrix} a_{11} & a_{12} & -- & a_{1n} & -1 & 0 & -- & 0 \\ a_{21} & a_{22} & -- & a_{2n} & 0 & -1 & -- & 0 \\ \text{---} & \text{---} & & & & & & \\ a_{n1} & a_{n2} & -- & a_{nn} & 0 & 0 & -- & -1 \\ c_{11} & c_{12} & -- & c_{1n} & 0 & 0 & -- & 0 \\ c_{21} & c_{22} & -- & a_{22} & 0 & 0 & -- & 0 \\ \text{---} & \text{---} & & & & & & \\ c_{n1} & c_{n2} & -- & c_{nn} & 0 & 0 & -- & 0 \end{vmatrix}$$

$$c_{ij} = [BA]_{ij}$$

or,

$$|N| = \begin{vmatrix} A & -\mathrm{I} \\ C & 0 \end{vmatrix}$$

whose value, by the previous geometrical argument should be $|C|$. That this is actually true is obtained at once from Laplace's development. Hence

$$|BA| = |B|\,|A|$$

since scalars commute and reversing the roles of B and A, we have at once

$$|AB| = |B|\,|A|$$

This fact can be seen quite clearly if one of the matrices, say A, has rows which are mutually perpendicular. Writing A and B as submatrices

$$AB = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ -- \\ \bar{X}_n \end{bmatrix} \begin{bmatrix} \bar{Y}_1, & \bar{Y}_2, & -- & \bar{Y}_n \end{bmatrix}$$

$$= \begin{bmatrix} \bar{X}_1\bar{Y}_1 & \bar{X}_1\bar{Y}_2 & -- & \bar{X}_1\bar{Y}_n \\ \bar{X}_2\bar{Y}_1 & \bar{X}_2\bar{Y}_2 & -- & \bar{X}_2\bar{Y}_n \\ -- & & & \\ \bar{X}_n\bar{Y}_1 & \bar{X}_n\bar{Y}_2 & -- & \bar{X}_n\bar{Y}_n \end{bmatrix}$$

$$|AB| = \begin{vmatrix} |\bar{X}_1||\bar{Y}_1| \cos \angle \bar{X}_1\bar{Y}_1 & |\bar{X}_1||\bar{Y}_2| \cos \angle \bar{X}_1\bar{Y}_2 & --- \\ |\bar{X}_2||\bar{Y}_1| \cos \angle \bar{X}_2\bar{Y}_1 & |\bar{X}_2||\bar{Y}_2| \cos \angle \bar{X}_2\bar{Y}_2 & --- \\ --- & & \\ |\bar{X}_n||\bar{Y}_1| \cos \angle \bar{X}_n\bar{Y}_1 & |\bar{X}_n||\bar{Y}_2| \cos \angle \bar{X}_n\bar{Y}_2 & --- \end{vmatrix}$$

$$= |\bar{X}_1||\bar{X}_2| -- |\bar{X}_n| \begin{vmatrix} |\bar{Y}_1| \cos \angle \bar{X}_1\bar{Y}_1 & |\bar{Y}_2| \cos \angle \bar{X}_1\bar{Y}_2 & --- \\ |\bar{Y}_1| \cos \angle \bar{X}_2\bar{Y}_1 & |\bar{Y}_2| \cos \angle \bar{X}_2\bar{Y}_2 & --- \\ --- & & \\ |\bar{Y}_1| \cos \angle \bar{X}_n\bar{Y}_1 & |\bar{Y}_2| \cos \angle \bar{X}_n\bar{Y}_2 & --- \end{vmatrix}$$

Now, since the rows of A are perpendicular vectors, $|A| = |\bar{X}_1||\bar{X}_2| \cdots |\bar{X}_n|$ $|A|$ thus being the volume of a rectangular parallelepiped, while $|\bar{Y}_i| \cos \angle \bar{X}_j \bar{Y}_i$ gives the coordinates of $\bar{Y}_i$; in terms of a new axis system parallel to the $\bar{X}_j$, so that the determinant in the expression above is just $|B|$. Hence $|AB| = |A||B|$. In other words, a matrix which multiplies a set of perpendicular vectors magnifies the volume which they span by a factor equal to its determinant.

It is evident geometrically that if there is a parallelepiped whose edges are unit vectors that the parallelepiped will assume its maximum volume $+1$ or its minimum volume -1 when each of the vectors is perpendicular to the remainder, and one has a hypercube. This fact may be derived from the formula for a determinant by considering the expression

$$\frac{\varphi^2(\bar{X}_1+\delta\bar{X}_1, \bar{X}_{-2}+\delta\bar{X}_{-2}, \ldots, \bar{X}_{-n}+\delta\bar{X}_n)}{(\bar{X}_1+\delta\bar{X}_1)(X+\delta X_1)(\bar{X}_2+\delta\bar{X}_2)(X_{-2}+\delta X_2)\cdots}$$

and setting the collection of first order differentials equal to zero, to find the stationary values of the determinant φ, and hence the maxima and minima. The conditions so derived will be necessary and not sufficient; the sufficiency arguments can be supplied by using the fact that $\varphi(\bar{X}_i)$ is continuous and defined over a bounded and closed region $\{X_i\}$.

$$\frac{\varphi^2(\bar{X}_i+\delta\bar{X}_{-i})}{\prod(\bar{X}_i+\delta\bar{X}_i)(X_i+\delta X_i)} = \frac{\varphi^2(\bar{X}_{-i}+\delta\bar{X}_{-i})}{\prod \bar{X}_i X_i + \sum_{i}\prod_{i\neq j} \bar{X}_y X_i\,\delta(\bar{X}_i X_j) + O(\delta X^2)}$$

where $O(\delta X^2)$ represents terms quadratic and higher in the δx_{ij}'s.

$$= \frac{\{\varphi(\bar{X}_i)+\sum\varphi(\bar{X}_{-i,ba,}\delta\bar{X}_i \ldots \bar{X}_n)+O(\delta X^2)\}^2}{\prod\bar{X}_{-i}X_i\left(1+\sum\frac{\delta(\bar{X}_i X_i)}{\bar{X}_i X_{-i}}+O(\delta X^2)\right)}$$

$$= \frac{\varphi(\bar{X}_i)^2 + 2\varphi(\bar{X}_i)\sum \varphi(\bar{X}_i \,--, \delta\bar{X}_i, -- X_n) + O(\delta^2)}{\prod \bar{X}_i X_i \left\{1 + \sum \frac{\delta(\bar{X}_i X_i)}{X_i X} + O(\delta^2)\right\}}$$

$$= \frac{1}{\prod \bar{X}_i X_{-i}} \left\{ \varphi(X_i)^2 + 2\varphi(\bar{X}_i)\sum_i \varphi(\bar{X}_{-1} \cdots, \delta\bar{X}_{-i} \cdots \bar{X}_n) + O(\delta^2)\right\} \times$$

$$\left\{1 - \sum_i \frac{\delta\bar{X}_i X_i}{\bar{X}_i X_{-i}} + O(\delta^2)\right\}$$

where the denominator has been expanded in geometric series.

When $\prod_i \bar{X}_i X_{-i} \neq 0$ the first order terms in δX can be equated to zero. Then

$$2\varphi(\bar{X})\sum_i \varphi(\bar{X}_{-1}, -- \delta\bar{X}_i -- X_n) = \varphi(\bar{X})^2 \sum_i \frac{\delta\bar{X}_i X_{-i}}{\bar{X}_{-i} X_i}$$

which is an expression which must hold for each δX_i. Then, setting all $\delta X_i = 0$ except for δX_j, and recalling that $\varphi(--- 0 ---) = 0$ and assuming that $\varphi \neq 0$

$$\varphi(\bar{X}_{-1}, -- \delta\bar{X}_{-j}, -- X_n) = \varphi(\bar{X}) \frac{\delta|X_j|^2}{2|X_j|^2}$$

the choice of the components of $\delta\bar{X}_j$ is still arbitrary, so if $[\delta\bar{X}_{-j}]_k = \delta x_{jk}$, set all $\delta x_{jk} = 0$ except δx_{jl}. Then, using the Laplace expansion:

$$\delta x_{jl}\,(\text{cofactor of } x_{jl}) = \varphi(\bar{X}_{-}) \frac{\delta x_{jl}}{|X_{-j}|^2}\, x_{jl}$$

if $\mathcal{X}_{jl} =$ cofactor of x_{jl}

$$\mathcal{X}_{jl} = \frac{\varphi(\bar{X}_j)}{|X_j|^2}\, x_{jl} \qquad \text{for each } j, l.$$

Now call

$$\Xi = \begin{bmatrix} x_{11} & x_{12} & -- & x_{1n} \\ x_{21} & x_{22} & -- & x_{2n} \\ -- & - & & \\ x_{n1} & x_{n2} & -- & x_{nn} \end{bmatrix}$$

$$\Upsilon = \begin{bmatrix} \chi_{11} & \chi_{12} & -- & \chi_{1n} \\ \chi_{21} & \chi_{22} & -- & \chi_{1n} \\ -- & - & & \\ \chi_{n1} & \chi_{n2} & -- & \chi_{nn} \end{bmatrix}$$

Now, $\Delta = \Xi \tilde{\Upsilon}$ has the form

$$\Delta = \begin{bmatrix} \varphi & 0 & -- & 0 \\ 0 & \varphi & -- & 0 \\ -- & - & & \\ 0 & 0 & -- & \varphi \end{bmatrix}$$

since

$$[\Delta]_{ij} = \sum_k [\Xi]_{ik} [\tilde{\Upsilon}]_{kl}$$

$$= \sum_k x_{ik} \chi_{jk}$$

$$= \varphi \delta_{ij}$$

where $i = j$ this expression is just Laplace's expansion for φ; where $i \neq j$ the result is a determinant whose i^{th} row is repeated twice, since χ_{jk} contains the i^{th} row, and whose j^{th} row is omitted; hence vanishes.

By the result of the last page, however,

$$\Upsilon = \begin{bmatrix} \frac{\varphi}{|X_1|^2} x_{11} & \frac{\varphi}{|X_1|^2} x_{12} & -- & \frac{\varphi}{|X_1|^2} x_{1n} \\ \frac{\varphi}{|X_2|^2} x_{21} & \frac{\varphi}{|X_2|^2} x_{22} & -- & \frac{\varphi}{|X_2|^2} x_{2n} \\ \frac{\varphi}{|X_n|^2} x_{n1} & --- & & \frac{\varphi}{|X_n|^2} x_{nn} \end{bmatrix}$$

and the product $\Xi \bar{Y} = \Delta$ now reads, in terms of the elements;

$$\varphi \delta_{ij} = \sum_k x_{ik} \frac{\varphi}{|X_j|^2} x_{jk}$$

so that

$$\bar{X}_i X_j = |X_j|^2 \delta_{ij}$$

a circumstance which requires the rows of the determinant φ to be perpendicular. If $|\bar{X}_i| = 1$, the resulting figure will be a cube of volume ± 1, which are the maximum and minimum values possible.

When two vectors satisfy the relation $\bar{X} Y = 0$, they are perpendicular. This property is often called orthogonality. Thus, for instance, the zero vector is orthogonal to every other vector. If a set of vectors all satisfy the relations $\bar{X}_i X_j = 0$, they are said to form an "orthogonal set" of vectors.

If a vector satisfies the condition $\bar{X} X = 1$, it is said to be "normal", or of unit length. Any vector which is not a null vector may be normalized by dividing it by the square root of its length.

If each vector of an orthogonal set is also normalized, the set is spoken of as an "orthonormal" set. The vector of such a set satisfy the relations

$$\bar{X}_i X_j = \delta_{ij}$$

A set of vectors is called "linearly dependent" if there can be found scalars α_i, not all of which are zero, such that

$$\alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n = 0$$

or in other words, if any one of them can be written as a sum of the others.

Any set containing the zero vector is automatically linearly dependent.

If no such constants can be found, the vectors are called linearly independent.

From the equation above there can be found a system of scalar equations; they are obtained from premultiplication of it in turn by $\bar{X}_1, \bar{X}_2$, etc.

$$\alpha_1 \bar{X}_1 X_1 + \alpha_2 \bar{X}_2 X_2 + \cdots + \alpha_n \bar{X}_1 X_n = 0$$

$$\alpha_1 \bar{X}_2 X_1 + \alpha_2 \bar{X}_2 X_2 + \cdots + \alpha_n \bar{X}_2 X_n = 0$$

$$- - -$$

$$\alpha_1 \bar{X}_n X_1 + \alpha_2 \bar{X}_n X_2 + \cdots \alpha_n \bar{X}_n X_n = 0$$

The problem is to choose α_i, not all zero, to satisfy these relations. In order for this to happen, the determinant

$$\begin{vmatrix} \bar{X}_1 X_1 & \bar{X}_1 X_2 & \cdots & \bar{X}_1 X_n \\ \bar{X}_2 X_2 & \bar{X}_2 X_2 & \cdots & \bar{X}_2 X_n \\ - - - & & & \\ \bar{X}_n X_1 & \bar{X}_n X_2 & \cdots & \bar{X}_n X_n \end{vmatrix}$$

must vanish. It is called the Gramm determinant, and its vanishment is the necessary and sufficient condition for linear dependence.

It can be factored, so that

$$\left| \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \bar{X}_n \end{bmatrix} [X_1, X_2 \cdots X_n] \right| = 0$$

or, if

$$M = [X_1, X_2 \cdots X_n]$$

By the rule for multiplying determinants

$$|M|\,|\bar{M}| = 0$$

whence
$$|M| = 0$$

which gives the geometric condition that a set of vectors are linearly dependent in n-space when the volume of the parallelepiped which they span vanishes, an altogether reasonable conclusion. In fact, if the vectors are normalized, this determinant, in giving the volume of their parallelepiped gives some idea of how "flat" the space they span is, since the smaller its volume, the flatter their parallelepiped, until its vanishing requires them all to lie in the same hyperplane.

Throughout this discussion it has been assumed that the volume of a parallelepiped is given by the determinant of its edges; in fact, it was in this manner that the formula for a determinant was attained. It is possible to prove this contention rather than relying upon intuitive geometrical arguments. The proof is contained as a corollary to the following theorem:

Let there be k vectors, $X_1, X_2, \ldots X_k$. If M is the matrix

$$M = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \cdots \\ \bar{X}_n \end{bmatrix}$$

the determinant $|M\bar{M}|$ gives the square of the volume of the parallelepiped which they form.

The proof is by induction. For $k=1$, the determinant gives $\bar{X}_1 X$, the square of the length of the vector, so that the theorem is in fact true for $k=1$.

Assume the theorem for $k=m$. Then for $k=m+1$ the determinant is

$$\begin{matrix} \bar{X}_1 X_1 & \bar{X}_2 X_2 & \cdots & \bar{X}_1 X_{m+1} \\ \bar{X}_2 X_1 & \bar{X}_2 X_2 & \cdots & \bar{X}_2 X_{m+1} \\ \cdots \\ \bar{X}_{m+1} X_1 & \bar{X}_{m+2} X_2 & \cdots & \bar{X}_{m+1} X_{m+1} \end{matrix}$$

Now write

$$X_{m+1} = \sum_{i=1}^{m} \alpha_i X_i + X^{\perp}$$

where $\bar{X}_i X^{\perp} = 0$. The α's may be found by solving the equations

$$\bar{X}_j X_{m+1} = \sum \alpha_i \bar{X}_j X_i \qquad j = 1, \ldots, m$$

for α_i $i = 1, \ldots, m$. The solution is possible unless $|M_m| = 0$. The determinant then becomes

$$|M_{m+1}| = \begin{vmatrix} \bar{X}_1 X_1 & \bar{X}_1 X_2 & \cdots & \sum_i \alpha_i \bar{X}_1 X_i \\ \bar{X}_2 X_1 & \bar{X}_2 X_2 & \cdots & \sum_i \alpha_i \bar{X}_2 X_i \\ \overline{} & \overline{} & \overline{} & \\ \sum \alpha_i \bar{X}_i X_1 & \sum \alpha_i \bar{X}_i X_2 & \cdots & \sum\sum \alpha_i \alpha_j \bar{X}_i X_j + \bar{X}^{\perp} X^{\perp} \end{vmatrix}$$

$$= \begin{vmatrix} \bar{X}_1 X_1 & \bar{X}_1 X_2 & \cdots & \sum \alpha_i \bar{X}_1 X_i \\ \bar{X}_2 X_1 & \bar{X}_2 X_2 & \cdots & \sum \alpha_i \bar{X}_2 X_i \\ \overline{} & \overline{} & \overline{} & \\ \sum \alpha_i \bar{X}_i X_1 & \sum \alpha_i \bar{X}_i X_2 & \cdots & \sum_i \sum_j \alpha_i \alpha_j \bar{X}_i X_j \end{vmatrix} + \begin{vmatrix} |M_m| & \sim \\ & \sim \\ & \sim \\ 0 \quad 0 \quad \cdots & \bar{X}^{\perp} X^{\perp} \end{vmatrix}$$

$$\begin{vmatrix} & & \\ & --- & \\ & & \end{vmatrix} + \bar{X}^{\perp} X^{\perp} |M_m|$$

the first determinant is zero, since the last row is line dependent upon the previous rows, since it was deliberately taken to be part of X_{m+1} lying in their plane, while the second term is of the form:

$$(\text{altitude})^2 \quad (\text{base})^2 = (\text{volume})^2$$

by the induction hypothesis.

The case $|M_m| = 0$ can perhaps be dealt with by renumbering the vectors, the order in which they appear being immaterial. If this is impossible, then each minor of the last row is zero, and Laplace's expansion gives a value of zero to the entire determinant. This is satisfactory since it assigns zero volume to a figure whose base has zero area.

If a set of n linearly independent vectors do not form an orthonormal set, it may be possible to define a "biorthonormal" set. This is a set of 2 n vectors, X_i^{I} and X_i^{II} such that

$$\bar{X}_i^{I} X_j^{II} = \delta_{ij}$$

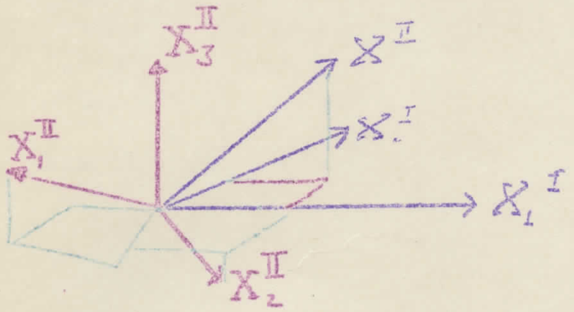
this expression is a representative element from the matrix equation

$$M \bar{N} = \mathrm{I}$$

where

$$M = \begin{bmatrix} X_1^{I} \\ X_2^{I} \\ \cdots \\ X_n^{I} \end{bmatrix} \qquad N = \begin{bmatrix} \bar{X}_1^{II} \\ \bar{X}_2^{II} \\ \cdots \\ \bar{X}_n^{II} \end{bmatrix}$$

which is capable of fulfillment when $|M| \neq 0$, for then $\bar{N} = M^{-1}$ exists. This is possible when the vectors are linearly independent, for then the determinant of their matrix is non-vanishing. Such a set of vectors is useful,



for instance, in dealing with an oblique coordinate system, in which the X^{I} are called the coordinate vectors and the X^{II} a "reciprocal set", having the property that each vector of the reciprocal set is perpendicular to the remaining coordinate vectors. While additions may be done with the coordinate vectors, for instance, the reciprocal set is required for a convenient discussion of projections. In fact, if one has

$$X = \sum \alpha_i X_i^{I}$$

$$\alpha_i = \bar{X}_i^{II} X$$

so that the reciprocal set are needed to isolate components.

From the above

$$X = \sum X_i^{I} (\bar{X}_i^{II} X)$$

$$= \sum_i (X_i^{I} \bar{X}_i^{II}) X$$

so that

$$\mathbb{I} = \sum_i X_i^{I} \bar{X}_i^{II}$$

a condition on a coordinate system that might be described as "completemess", since then any vector in the space in question can be written in terms of the

X_i^{I} as coordinates. Since

$$\sum \bar{X} X_{\cdot i}^{I} \bar{X}_i^{II} = \bar{X}$$

$$\bar{X} = \sum (\bar{X} \cdot X_{\cdot i}^{I}) \bar{X}_i^{II}$$

it is seen that the two systems play equivalent roles, and that either may be regarded as the coordinate system with the other forming the reciprocal system.

MATRIX FUNCTIONS

From a given matrix others may be found by various processes. Square matrices, for instance, may be raised to powers, added, and in general combined as though they were scalars, in building up functions; it being neccessary to bear in mind their non-commitativity.

A single square matrix may be raised to powers; the integral powers are defined by

$$\underbrace{A \cdot A \cdot A \cdot \text{----} A}_{n \text{ factors}} = A^n$$

while if $B^m = A$, one speaks of B as the mth root of A; $A^{1/m}$. Since the commutative law fails one cannot expect that there will be only m mth roots of a matrix; in fact, there are many more in some cases.

If A has an inverse A^{-1}, the negative powers are defined by

$$A^{-n} = (A^{-1})^n$$

The zero power of a non-singular matrix is the unit matrix:

$$A^0 = I \qquad |A| \neq 0$$

since the law of exponents holds, from the definitions, so that

$$AA^{-1} = A^0 = I$$

If the matrix is singular, forming the zeroth power produces a situation analogous to that arising when one tries to find 0/0.

Polynomials in a matrix may be formed, and the coefficients may be either scalars or matrices, although one usually means a polynomial with scalar coefficients in referring to a matrix polynomial

$$P(A) - \sum_{N=1}^{N} \alpha_n A^n$$

By taking the limit of a polynomial with a large number of terms it is possible to obtain power series in matrices. The finite geometric series may be summed by the same technique used with scalar series for instance. Let

$$\begin{aligned} S = \sum_0^n A^i &= \mathbb{I} + A + A^2 + --- + A^n \\ &= \mathbb{I} + A(\mathbb{I} + A + A^2 + --- + A^{n-1}) \\ &= \mathbb{I} + A(\mathbb{I} + A + --- + A^n) - A^{n+1} \\ &= \mathbb{I} + AS - A^{n+1} \end{aligned}$$

$$S(\mathbb{I} - A) = \mathbb{I} - A^{n+1}$$

$$S = \{\mathbb{I} - A^{n+1}\}\{\mathbb{I} - A\}^{-1}$$

If $\lim_{n\to\infty} A^n = 0$, which means that $\lim_{n\to\infty} [A^n]_{ij} \to 0$

$$S = \{\mathbb{I} - A\}^{-1}$$

$$\therefore \sum_0^\infty A^n = \{\mathbb{I} - A\}^{-1}$$

The matrix exponential is defined as

$$e^A = \mathbb{I} + A + \frac{1}{2!}A^2 + --- = \sum_0^\infty \frac{A^n}{n!}$$

It converges for each square matrix A of finite order. To show this let the absolute value of the largest element of A by μ . Furthermore define matrix inequality to hold for each element

A < B when $[A]_{ij} < [B]_{ij}$

then

$$A \leq \begin{bmatrix} \mu & \mu & -- & \mu \\ - & - & & \\ \mu & \mu & -- & \mu \end{bmatrix}$$

$$\leq \mu \begin{bmatrix} 1 & 1 & -- & 1 \\ - & - & & \\ 1 & 1 & -- & 1 \end{bmatrix}$$

call $$\Xi = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

$$\Xi^2 = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} n & n & \cdots & n \\ \vdots & \vdots & & \vdots \\ n & n & \cdots & n \end{bmatrix} = n\,\Xi$$

so that $$\Xi^k = n^{k-1}\,\Xi$$

This means that, in absolute value, each term of the series is less than

$$\frac{\mu^k n^{k-1}}{k!}\,\Xi$$

so that, in absolute value, each element of e^A is less than

$$1 + \sum_1^\infty \frac{\mu^k n^{k-1}}{k!} = \frac{1}{n}\left(e^{\mu n} - 1\right) + 1$$

which is surely finite for each finite μ and n. Hence the series converges absolutely for each square matrix.

Unlike the scalar exponential, the rule for adding logarithms fails unless the logarithms commute. To see this, write

$$\begin{aligned} e^A e^B = {} & \mathbb{I} + B + \frac{B^2}{2!} + \cdots \\ \times {} & \mathbb{I} + A + \frac{A^2}{2!} + \cdots \end{aligned}$$

$$\begin{aligned} = {} & \mathbb{I} + B + \tfrac{1}{2!} B^2 + \cdots \\ & + A + AB + \tfrac{1}{2!} AB^2 + \cdots \\ & + \tfrac{1}{2!} A^2 + \tfrac{1}{2!} A^2 B + \tfrac{1}{2!\,2!} A^2 B^2 + \cdots \\ & + \cdots \end{aligned}$$

Summing this by diagonals,

$$= \mathbb{I} + (A+B) + \frac{1}{2!}\left(A^2 + 2AB + B^2\right) + \cdots$$

If AB = BA, this is

$$= \mathbb{I} + (A+B) + \frac{1}{2!}\left(A^2 + AB + BA + B^2\right) + \frac{1}{3!}\left(A^3 + A^2B + ABA + BA^2 + \cdots\right) + \cdots$$

$$= \mathbb{I} + (A+B) + \frac{1}{2!}(A+B)^2 + \cdots$$

$$= e^{(A+B)}$$

Now, A and -A commute so that

$$e^{A}e^{-A} = e^{\mathbb{O}} = \mathbb{I}$$

a fact which could have been verified directly. However, it shows that

$$(e^{A})^{-1} = e^{-A}$$

Logarithms may be added when they commute. If they are infinitesimal they may still be added, to first order. Thus

$$\begin{aligned} e^{\varepsilon A}e^{\varepsilon B} &= (\mathbb{I} + \varepsilon A + \cdots)(\mathbb{I} + \varepsilon B + \cdots) \\ &= (\mathbb{I} + \varepsilon(A+B) + \varepsilon^2 A\,B + \cdots \\ &= \mathbb{I} + \varepsilon\,(A+B) + \cdots \\ &= e^{\varepsilon\,(A+B)} \end{aligned}$$

If the elements of a matrix are functions of a scalar parameter, the matrix may be integrated and differentiated with respect to this parameter. The derivative is defined, in the usual manner, as

$$\lim_{\Delta t \to 0} \frac{A(t+\Delta t) - A(t)}{\Delta t} = \frac{d\,A(t)}{dt}$$

each of these operations may be performed directly upon the individual elements, giving

$$\left[\lim_{\Delta t \to 0} \frac{[A(t+\Delta t)]_{ij} - [A(t)]_{ij}}{\Delta t}\right] = \frac{dA}{dt}$$

so that $$\left[\frac{dA}{dt}\right]_{ij} = \frac{d\,[A]_{ij}}{dt}$$

and a matrix is differentiated by differentiating each of its elements.

The derivatives of functions of a matrix may also be taken, with formulas closely resembling the scalar formulae, except where the failure of the commutative law must be taken into account.

$$1. \quad \frac{d}{dt}\{A+B\} = \frac{dA}{dt} + \frac{dB}{dt}$$

since:
$$\left[\frac{d(A+B)}{dt}\right]_{ij} = \frac{d}{dt}[(A+B)]_{ij}$$
$$= \frac{d}{dt}\{[A]_{ij} + [B]_{ij}\}$$
$$= \frac{d[A]_{ij}}{dt} + \frac{d[B]_{ij}}{dt}$$
$$= \left[\frac{dA}{dt} + \frac{dB}{dt}\right]_{ij}$$

$$2. \quad \frac{dAB}{dt} = \frac{dA}{dt}B + A\frac{dB}{dt}$$

Since:
$$\left[\frac{dAB}{dt}\right]_{ij} = \frac{d}{dt}[AB]_{ij}$$
$$= \frac{d}{dt}\left\{\sum_k [A]_{ik}[B]_{kj}\right\}$$
$$= \sum_k \frac{d[A]_{ik}}{dt}[B]_{kj} + \sum_k [A]_{ik}\frac{d[B]_{kj}}{dt}$$
$$= \left[\frac{dA}{dt}B\right]_{ij} + \left[A\frac{dB}{dt}\right]_{ij}$$
$$= \left[\frac{dA}{dt}B + A\frac{dB}{dt}\right]_{ij}$$

while a similar rule holds for products of more than two matrices.

In differentiating powers, say A^3, one has
$$\frac{dA^3}{dt} = A^2\frac{dA}{dt} + A\frac{dA}{dt}A + \frac{dA}{dt}A^2$$

A and dA/dt do not usually commute. If A commutes with itself for all values of its parameters
$$A(t_1)A(t_2) = A(t_2)A(t_1)$$
the matrix and its derivative will commute. The reason for this is that the

derivative involves the difference of such matrices. If A commutes with them it will commute with their difference and hence with the derivative.

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

proof: $$AA^{-1} = I$$

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

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

$$\frac{dA^{-1}}{dt} = -A^{-1}\frac{dA}{dt}A^{-1} \qquad \text{qed.}$$

1. $$\frac{de^A}{dt} = e^A\frac{dA}{dt} = \frac{dA}{dt}e^A$$

$$\frac{de^A}{dt} = \frac{d}{dt}\left\{I + A + A^2/2! + \cdots\right\}$$

$$= \frac{dA}{dt} + \frac{1}{2!}\left(A\frac{dA}{dt} + \frac{dA}{dt}A\right\} + \cdots$$

if dA/dt commutes with A, the terms in $\frac{dA}{dt}$ may be factored out,

$$= \frac{dA}{dt}\left\{I + A + \frac{1}{2!}A^2 + \cdots\right\}$$

$$= \left\{I + A + \frac{1}{2!}A^2 + \cdots\right\}\frac{dA}{dt}$$

$$= e^A\frac{dA}{dt} = \frac{dA}{dt}e^A \qquad \text{qed.}$$

To integrate a matrix with respect to a parameter, write the integral as a limit of sum

$$\int_{t_0}^{t} A(\alpha)\,d\sigma = \lim_{\Delta t \to 0} \sum A(t_i)\Delta t_i$$

which, being composed of a limit and a sum, is an operation upon the individual elements of the matrix

$$= \left[\lim_{\Delta t \to 0} \sum [A(t_i)]_{lm} \Delta t\right]$$

so that

$$\left[\int_{t_0}^{t} A(\sigma)\, d\sigma\right]_{ij} = \int_{t_0}^{t} [A(\sigma)]_{ij}\, d\sigma$$

and a matrix is integrated by integrating the individual elements.

With the introduction of derivatives one must contend with the possibility of differential equations. Just as matrix methods were used to solve linear equations in n unknowns, so they may be used to solve systems of linear differential equations, such as

$$\frac{dx_1(t)}{dt} = f_{11}(t)x_1(t) + f_{12}(t)x_2(t) + \cdots + f_{1n}(t)x_n(t)$$

$$\frac{dx_2(t)}{dt} = f_{21}(t)x_1(t) + f_{22}(t)x_2(t) + \cdots + f_{2n}(t)x_n(t)$$

$$-\ -\ -$$

$$\frac{dx_n(t)}{dt} = f_{n1}(t)x_1(t) + f_{n2}(t)x_2(t) + \cdots + f_{nn}(t)x_n(t)$$

$$\begin{bmatrix} \frac{dx_1(t)}{dt} \\ \frac{dx_2(t)}{dt} \\ -\ - \\ \frac{dx_n(t)}{dt} \end{bmatrix} = \begin{bmatrix} f_{11} & f_{12} & -- & f_{1n} \\ f_{21} & f_{22} & -- & f_{2n} \\ -- & & & \\ f_{n1} & f_{n2} & -- & f_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ -- \\ x_n \end{bmatrix}$$

Discretize the equation $\dfrac{d\mathbf{X}}{dt} = \mathbf{F}\mathbf{X}$

$$\frac{1}{\Delta t}\left\{\mathbf{X}(t)\Delta t) - \mathbf{X}(t)\right\} = \mathbf{F}(t)\mathbf{X}(t)$$

$$\mathbf{X}(t+\Delta t) = \left\{\mathbf{I} + \Delta t\, \mathbf{F}(t)\right\}\mathbf{X}(t)$$

$$= e^{\Delta t\, F(t)}\, \mathbf{X}(t)$$

to first order in Δt. If $\sum \Delta t_i = \tau$, we have

$$X(t+\tau) = e^{\Delta t_n F(t_n)} e^{\Delta t_{n-1} F(t_{n-1})} \cdots e^{\Delta t_1 F(t_1)} X(t)$$

If all the exponents commute,

$$= e^{F(t_n)\Delta t_n + F(t_{n-1})\Delta t_{n-1} + \cdots + F(t_1)\Delta t_1} X(t)$$

taking the limit,

$$X(t+\tau) = \lim_{\Delta t_n \to 0} e^{\sum F(t_i)\Delta t_i} X(t)$$

$$X(t+\tau) = e^{\int_t^{t+\tau} F(\sigma)\, d\sigma} X(t)$$

this assumes a particularly simple form when F is a matrix of constants:

$$\int_t^{t+\tau} F(\sigma)\, d\sigma = F\tau$$

$$X(t+\tau) = e^{F\tau} X(t)$$

whose validity may be checked directly, and which resembles the solution to the corresponding scalar equation.

If the matrix F does not commute with itself for all values of the parameter, the solution of the equation is more difficult. Recalling

$$X(t+\Delta t) = \{I + \Delta t\, F(t)\} X(t)$$

and writing

$$X(t+\tau) = \{I + \Delta t_n F(t_n)\}\{I + \Delta t_{n-1} F(t_{n-1})\} \cdots \{I + \Delta t F(t)\} X(t)$$

which is still correct to the first order in Δt. Calling

$$\Omega_n = \{I + \Delta t_n F(t_n)\}\{I + \Delta t_{n-1} F(t_{n-1})\} \cdots \{I + \Delta t_1 F(t_1)\}$$

we have

$$\Omega_n = I + \sum_{i=1}^{N} F(t_i)\Delta t_i + \sum_{i=1}^{N} \sum_{j<i} F(t_i) F(t_j) \Delta t_i \Delta t_j$$

$$+ \sum_{i=1}^{N} \sum_{j<i} \sum_{k<j} F(t_i) F(t_j) F(t_k) \Delta t_i \Delta t_j \delta t_k + \cdots$$

the general term of this series is the r^{th};

$$\sum_{i=1}^{n} \sum_{j<i} \sum_{k<j} \cdots \sum_{l<m} F(t_i) F(t_j) \cdots F(t_m) \Delta t_i \Delta t_j \cdots \Delta t_m$$

there are $N(N-1)\cdots(N-r+1)/r!$ terms included. In such a multiple sum, which is of the order

$$\frac{N^R}{R!} + O(N^{R-1})$$

each summand has the order of magnitude Δt^R, so that the whole sum is of the order of magnitude

$$\frac{N^R \Delta t^R}{R!} + O \frac{\Delta t}{R!}$$

but $\tau \approx N \Delta t$, so that they are of relative size

$$\frac{\tau^R}{R!} + O\left(\frac{\Delta t}{R!}\right)$$

and hence the series for Ω converges as well as the exponential series. Taking $\lim_{n \to \infty} \Omega_n$, term by term, there arise expressions

$$\lim_{n\to\infty} \sum_{i=1}^{N} \sum_{j<i} \cdots \sum_{l<m} F(t_i)\, F(t_j) \cdots F(t_m) \Delta t_i \Delta t_j \cdots \Delta t_m$$

$$= \int_t^{t+\tau} \int_t^{t+\sigma_1} \cdots \int_t^{t+\sigma_{k-1}} F(\sigma_1) F(\sigma_2) \cdots F(\sigma_k) d\sigma_k\, d\sigma_{k-1} \cdots d\sigma_2\, d\sigma_1$$

so that

$$\Omega = I + \int_t^{t+\sigma} F(\sigma)\, d\sigma + \int_t^{t+\sigma} \int_t^{\sigma_1+t} F(\sigma_1) F(\sigma_2)\, d\sigma_2\, d\sigma_1$$

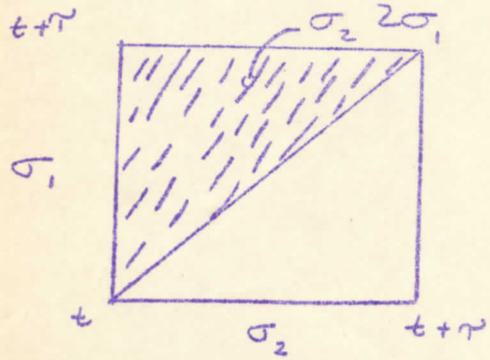
$$+ \cdots$$

$$X(t+\tau) = \Omega X(t)$$

that this expression actually satisfies the differential equation may be checked by term by term differentiation of the series.

In the case that all the F's commute, the range of integration need not be restricted to preserve the order of the factors. Then one has, for instance

$$\int_t^{t+\tau} \int_t^{t+\sigma_1} F(\sigma_1)F(\sigma_2)\,d\sigma_2\,d\sigma_1 = \frac{1}{2} \int_t^{t+\tau} \int_t^{t+\tau} F(\sigma_1)F(\sigma_2)\,d\sigma_1\,d\sigma_2$$



$$= \frac{1}{2}\left\{ \int_t^{t+\tau} F(\sigma)\,d\sigma \right\}^2$$

the region where $\sigma_2 < \sigma_1$, is just half the area over which the double integration of the second integral is extended, giving the factor 1/2.

Similarly the other integrals go over into

$$\frac{1}{k!}\left\{ \int_t^{t+\tau} F(\sigma)\,d\sigma \right\}^k$$

which gives the previous result.

EIGENVECTORS

It was shown that when multiplied by a matrix, a given vector was changed into another. The elements of the matrix give the dependence of the elements of the new vector upon those of the old. One important class of vectors, which depend upon particular matrices, are the vectors whose direction remains unchanged by matrix multiplication. Such vectors are called eigenvectors; and since they are different for different matrices they are spoken of as belonging to a certain matrix.

They satisfy the matrix equation

$$AX = \lambda X$$

where λ is a scalar, called an eigenvalue, and reflects the fact that the length of the vector may change even though its direction is held fixed. Since different eigenvectors will be changed in length by different amounts in general, it is customary to refer to eigenvalues belonging to a certain eigenvector, since in this way they serve to distinguish the different eigenvectors from one another.

A trivial solution to the eigenvector equation is $\mathbb{X} = \mathbb{O}$, the zero vector. It is disallowed as an eigenvector. It should be noted that this is not the case with null vectors, whose lengths are merely zero, without being zero themselves. If they arise they are quite legitimate eigenvectors.

The equation

$$A\mathbb{X} = \lambda\mathbb{X}$$

may be written

$$\{A - \lambda\mathbb{I}\}\mathbb{X} = 0$$

Suppose that $\{A - \lambda\mathbb{I}\}$ had an inverse. Then by multiplying by it from the left,

$$\{A - \lambda\mathbb{I}\}^{-1}\{A - \lambda\mathbb{I}\}\mathbb{X} = \{A - \lambda\mathbb{I}^{-1}\}\mathbb{O}$$

$$\mathbb{X} = \mathbb{O}$$

which has been forbidden. Thus it follows that $\{A - \lambda\mathbb{I}\}$ is singular, hence that its determinant

$$|A - \lambda\mathbb{I}|$$

must vanish

$$\begin{vmatrix} a_{11} - \lambda & a_{12} & --- & a_{1n} \\ a_{21} & a_{22} - \lambda & --- & a_{2n} \\ --- & & & \\ a_{n1} & a_{n2} & --- & a_{nn} - \lambda \end{vmatrix} = 0$$

It may be expanded in powers of λ to obtain

$$(-\lambda)^n + (-\lambda)^{n-1} \sum_i a_{ii} + \frac{(-\lambda)^{n-2}}{2!} \sum_i \sum_j \begin{vmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{vmatrix}$$

$$+ --- + \begin{vmatrix} a_{11} & a_{12} & -- & a_{1n} \\ a_{21} & a_{22} & -- & a_{2n} \\ -- & & & \\ a_{n1} & a_{n2} & -- & a_{nn} \end{vmatrix} = 0$$

The first term is got by taking the term of the expansion containing all the diagonal elements; the second by considering the term with $(n-1)\ \lambda$'s, which can come only from the diagonal and one other term, which must also come from the diagonal, and not being a λ is an a_{ii}. Summing for each possible way to do this gives $\sum a_{ii}$ as the coefficient. The third term is obtained by multiplying $(n-2)\ \lambda$'s together, and two other terms. These must either sit on the diagonal, giving $a_{ii}\, a_{jj}$, or one may sit off, say a_{ji}, whence the other must be a_{ij}, giving $-a_{ij}\, a_{ji}$, the one interchange needed to put the second indices in order contributing the minus sign. These two terms may then be combined into the single term

$$\begin{vmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{vmatrix}$$

while the 1/2 reflects the fact that the summation should be $\sum_i \sum_{j<i}$ rather than $\sum_i \sum_j$

The other terms are formed in a similar manner, giving finally the term free of λ's, which is gotten from the expansion by taking the diagonal elements and multiplying the a's together, which is the same term as would occur in the expansion of $|A|$.

Since this is a polynomial equation in λ, it can have at most n roots, hence at most n distinct eigenvalues. This equation is called the characteristic equation; the polynomial the characteristic polynomial.

If the eigenvalues of a small matrix are known, the eigenvectors may often be obtained by simple manipulation. For example, consider the matrix

$$S' = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

its characteristic equation is

$$\begin{vmatrix} -\lambda & 1 & 0 \\ 0 & -\lambda & 1 \\ 1 & 0 & -\lambda \end{vmatrix} = (-\lambda)^3 + 1 = 0$$

$$\lambda^3 = 1$$

$$\lambda = 1, \omega, \omega^2$$

$$\text{with} \quad \omega = e^{i\frac{2\pi}{3}}$$

The eigenvector equation is, for $\lambda = 1$

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 1 \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

$$\begin{bmatrix} b \\ c \\ a \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

$a = b = c$

so that the normalized eigenvector belonging to $\lambda = 1$ is $\frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. From this it appears that the length of an eigenvector is completely indeterminant, since matrix multiplication is linear. eg

$$A(2\mathbf{x}) = \lambda(2\mathbf{x})$$

and $\quad A\mathbf{x} = \lambda\mathbf{x} \quad$ both hold.

The two other eigenvectors are readily determined to be:

$$\text{for } \lambda = \omega: \begin{bmatrix} 1/\sqrt{3} \\ \omega/\sqrt{3} \\ \omega^2/\sqrt{3} \end{bmatrix} \qquad \text{for } \lambda = \omega^2: \begin{bmatrix} 1/\sqrt{3} \\ \omega^2/\sqrt{3} \\ \omega/\sqrt{3} \end{bmatrix}$$

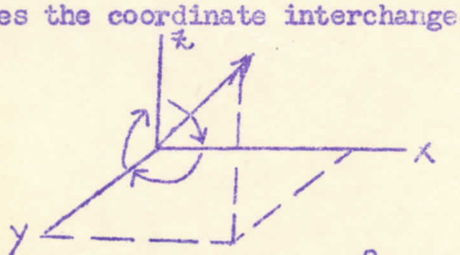
and it is of interest to note that these are null vectors:

$$|\mathbf{x}|^2 = \frac{1}{a} + \frac{\omega^2}{a} + \frac{\omega}{a} = \frac{1}{9}\left(1 + \left(\frac{1}{2} - \frac{i\sqrt{3}}{3}\right)^2 + \left(\frac{1}{2} + \frac{i\sqrt{3}}{3}\right)^2\right) = 0$$

The geometrical interpretation of the first eigenvector is obtained from noticing that since

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} y \\ z \\ x \end{bmatrix}$$

makes the coordinate interchanges:



$x \to y$
$y \to z$
$z \to x$

which corresponds to a 120° rotation about the axis $(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$
The eigenvectors belonging to $\lambda = 1$ are then the various vectors pointing along the axis of rotation; they are left unchanged by the rotation. The other two, being comples, do not have an interpretation in a real 3-dimensional space.

If two eigenvectors have the same eigenvalue, then any linear combination of them has the same eigenvalue.

Then $$A(a\mathfrak{X}+b\mathfrak{Y}) = Aa\mathfrak{X} + Ab\mathfrak{Y} = a(A\mathfrak{X})+b(A\mathfrak{Y})$$

$$= a\lambda\mathfrak{X} + b\lambda\mathfrak{Y} = \lambda(a\mathfrak{X})\, b\mathfrak{Y})$$

so that if $\mathfrak{X}$ and $\mathfrak{Y}$ are not parallel, each vector in the plane which they determine is an eigenvector with the same eigenvalue. An example of such a matrix is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

each vector in the $x-y$ plane has the eigenvalue 1.

The equation $A\mathfrak{X} = \lambda\mathfrak{X}$ defines only right eigenvectors. Left eigenvectors are defined by

$$\overline{\mathfrak{X}}A = \lambda\overline{\mathfrak{X}}$$

which leads to the same characteristic equation and thus the same eigenvalues as for the right eigenvectors.

Much information concerning the eigenvectors and the eigenvalues of a matrix may be deduced in a purely formal way, by manipulating their symmetry properties.

Theorem 1. All the eigenvalues of a hermitean matrix are purely real.

$$A\mathfrak{X} = \lambda\mathfrak{X} \qquad 1)$$

$$\overline{\mathfrak{X}}^*A\mathfrak{X} = \lambda\overline{\mathfrak{X}}^*\mathfrak{X} \qquad 2)$$

$$\overline{\mathfrak{X}}^*\overline{A}^* = \lambda^*\overline{\mathfrak{X}}^* \qquad -{}^*\ \text{eqn } 1)$$

$$\overline{\mathfrak{X}}^*\overline{A}^*\mathfrak{X} = \lambda^*\overline{\mathfrak{X}}^*\mathfrak{X} \qquad 4)$$

$$\lambda = \frac{\overline{\mathfrak{X}}^*A\mathfrak{X}}{\overline{\mathfrak{X}}^*\mathfrak{X}} \qquad \text{from } 2)$$

$$\lambda^* = \frac{\overline{\mathfrak{X}}^*\overline{A}^*\mathfrak{X}}{\overline{\mathfrak{X}}^*\mathfrak{X}}$$

since $\bar{A}^* = A$ by the hermitean property

$$\lambda = \lambda^*$$

qed

In a similar manner it is proven that

Theorem II. All the eigenvalues of an antihermitean matrix are purely imaginary.

Theorem III. Either the eigenvalues of a real symmetric matrix are real or they have null eigenvectors.

Theorem IV. Either the eigenvalues of an antisymmetric matrix are zero or they have null eigenvectors.

Theorem V. The left eigenvectors of a hermitean matrix are the complex conjugate transposes of the right eigenvectors.

$$AX = \lambda X$$
$$\bar{X}^* \bar{A}^* = \lambda^* \bar{X}^*$$
$$\lambda = \lambda^* \qquad A = \bar{A}^*$$
$$\therefore \bar{X}^* A = \lambda \bar{X}^*$$

qed

Theorem VI. The left eigenvectors and right eigenvectors of a matrix, which belong to different eigenvalues are orthogonal to one another in the hermitean sense $\bar{X}^* Y = 0$. This is sometimes expressed by saying that they are unitary to one another.

then
$$MX = \lambda_i X$$
$$\bar{Y}^* M = \mu \bar{Y}^*$$
$$\bar{Y}^* M X = \lambda \bar{Y}^* X$$
$$\bar{Y}^* M X = \mu \bar{Y}^* X$$
$$(\lambda - \mu) \bar{X}^* X = 0$$

$\lambda \neq \mu \quad \therefore \bar{Y}^* X = 0$

Information concerning functions of matrices may be obtained

Theorem VII. The eigenvalues of A^{-1} are the reciprocals of the eigenvalues of A, and have the same eigenvectors.

$$AX = \lambda X$$
$$\lambda^{-1} A^{-1} A X = \lambda^{-1} \lambda A^{-1} X$$
$$A^{-1} X = \lambda^{-1} X$$

Theorem VIII. The eigenvalues of A^2 are the square~~d~~ of the eigenvalues of A, and have the same eigenvector:

$$AX = \lambda X$$
$$A^2(X) = A(AX) = A\lambda X$$

A similar result holds for other powers, and polynomials, and functions expressible in convergent power series.

Theorem IX. For each matrix A, the eigenvalues of $\bar{A}^* A$ are real and non-negative.

They are real since $\bar{A}^* A$ is hermitean

$$\bar{A}^* A X = \lambda X$$
$$\bar{X}^* \bar{A}^* A X = \lambda \bar{X}^* X$$
$$\overline{AX}^* AX = \lambda(\bar{X}^* X)$$
$$\lambda = \frac{\overline{AX}^* AX}{\bar{X}^* X}$$

$$\bar{X}^* X > 0 \qquad \overline{AX}^* AX \geq 0$$

$$\therefore \lambda \geq 0$$

From Theorem VI it appears that if all the eigenvalues of a matrix are distinct, the left and right eigenvectors form a biorthogonal set, in the hermitean sense.

Since the vectors are determined at best by direction and not at all by length, it would be convenient to deal with a set of unit eigenvectors. This is still not possible, due to the possibility of the occurence of null vectors, since a null eigenvector cannot be normalized.

For the cases where some sort of normalization is possible, it is convenient to use a distinctive notation. Thus one writes

$$A|i\rangle = \lambda_i \, |i\rangle$$
$$\langle j|A = \lambda_j \langle j|$$

where $\overline{\langle j|}^{*} = |j\rangle$ only in special cases, such as when A is hermitean. These vectors can be distinguished from the ordinary bras and kets in that there is a number written in them, designating them as the i^{th} eigenbra or eigenket, while ordinary vectors have merely elements, as $|x_i\rangle$ with x_i written in them. Thus

$$\langle i|j\rangle = \delta_{ij}$$

is presumed to hold whenever such a choise is possible; the eigenvector then being taken as "normalized" by convention. With a hermitean matrix, for instance, this normalization is always possible.

If perchance $\langle i|i\rangle = 0$, it will be impossible to form a complete set of eigenvectors, since in a complete system

$$X = \alpha_1 |1\rangle + \alpha_2 |2\rangle + \cdots + \alpha_n |n\rangle \quad a$$
$$\langle i|X = a_i \langle i|i\rangle$$

but, it is possible to choose $\langle i|X \neq 0, \therefore \langle i|i\rangle \neq 0$, and the vectors could then be normalized to make

$$\langle i|i\rangle = 1$$, a contradiction

It should be noted that in a biorthonormal set one requires only

$$\langle i | i \rangle = 1$$

not that $|\langle i | \, | = 1$, as the condition for normalization.

Let us investigate the hypothesis that the eigenvectors of a certain matrix form a complete biorthonormal set. The condition for this is that

$$\Sigma \quad | i \rangle \langle i | = \underline{\underline{I}}$$

Such a sum is an inner product of the supervectors

$$P = [\, |1\rangle, \; |2\rangle, \; ---, \; |n\rangle \,]$$

$$P^{-1} = \begin{bmatrix} \langle 1 | \\ \langle 2 | \\ -- \\ \langle n | \end{bmatrix}$$

since $\quad P \, P^{-1} = \sum_{i=1}^{n} |i\rangle\langle i| = \underline{\underline{I}}$

But also

$$P^{-1} P = \begin{bmatrix} \langle 1|1\rangle & \langle 1|2\rangle & --- & \langle 1|n\rangle \\ \langle 2|1\rangle & \langle 2|2\rangle & --- & \langle 2|n\rangle \\ -- & & & \\ \langle n/1\rangle & \langle n|2\rangle & --- & \langle n/n\rangle \end{bmatrix} = \underline{\underline{I}}$$

so that $\langle i | y \rangle = \delta i j$ and that

$|P| \neq 0$, which is the true condition for completeness.

The hypothesis is not always true, as for instance in the case with the matrix

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

Its eigenvalues satisfy

$$\begin{vmatrix} 1-\lambda & 1 \\ 0 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 = 0$$

$$\lambda = 1, 1$$

so that there is a repeated eigenvalue.

$$[a, b] \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = [a, b]$$

$$[a, a+b] = [a, b]$$

$$\therefore a = 0, \quad b \text{ arbitrary}$$

$$\boxed{\langle 1| = [0, \ 1]}$$

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

$$\begin{bmatrix} \alpha+\beta \\ \beta \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

$$\beta = 0, \quad \alpha \text{ arbitrary}$$

$$\boxed{|1\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}}$$

so that, for one thing, there is only one eigenvector to span a two dimensional space, but also

$$\langle 1|1\rangle = [0, 1] \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 0$$

From the completeness hypothesis follows a theorem known as the "Spectral theorem"

$$A = \sum_{i=1}^{n} \lambda_i |i\rangle\langle i|$$

These are the spectral projections E_i.
NOTE: $E_i E_j = \begin{cases} 0 & \text{if } i \neq j \\ E_i & \text{if } i = j. \end{cases}$

Proof:

$$A|i\rangle = \lambda_i |i\rangle$$

$$A|i\rangle\langle i| = \lambda_i |i\rangle\langle i|$$

$$\sum_{i=1}^{n} A|i\rangle\langle i| = \sum_{i=1}^{n} \lambda_i |i\rangle\langle i|$$

$$A \sum_{i=1}^{n} |i\rangle\langle i| = A\mathbb{I} = \sum_{i=1}^{n} \lambda_i |i\rangle\langle i|$$

qed

Furthermore, considering the product $P^{-1}AP$

$$P^{-1}AP = \begin{bmatrix} \langle 1| \\ \langle 2| \\ \langle 3| \\ -- \\ \langle n| \end{bmatrix} A \left[|1\rangle \ |2\rangle \ \cdots \ |n\rangle \right]$$

$$\begin{bmatrix} \langle 1| \\ \langle 2| \\ -- \\ \langle n| \end{bmatrix} \left[\lambda_1 |1\rangle \quad \lambda_2 |2\rangle \ -- \ \lambda_n |n\rangle \right]$$

$$\begin{bmatrix} \lambda_1 \langle 1|1\rangle & \lambda_2 \langle 1|2\rangle & --- & \lambda_n \langle 1|n\rangle \\ \lambda_1 \langle 2|1\rangle & \lambda_2 \langle 2|2\rangle & --- & \lambda_n \langle 2|n\rangle \\ --- & & & \\ \lambda_1 \langle n|1\rangle & \lambda_2 \langle n|2\rangle & --- & \lambda_n \langle n|n\rangle \end{bmatrix}$$

$$\begin{bmatrix} \lambda_1 & 0 & -- & 0 \\ 0 & \lambda_2 & -- & 0 \\ -- & & & \\ 0 & 0 & -- & \lambda_n \end{bmatrix} = \Delta$$

Such a matrix, which has elements of the form $\lambda_i \delta_{ij}$, ie with elements only upon the main diagonal, is called a diagonal matrix. The transformation $P^{-1}AP$ is called a similarity transformation, and if $P^{-1}AP = \Delta$, a diagonal matrix, the process is called diagonalization.

The diagonalization process facilitates the proof of another important theorem: namely Sylvester's theorem:

if $\sum |i\rangle\langle i| = I$ and $f(\lambda_j)$ converges for each λ_j; as a Taylor's series, then

$$f(A) = \sum_{j=1}^{n} f(\lambda_j)\, |j\rangle\langle j|$$

Proof: if P diagonalizes a it also diagonalizes A^2 for

$$P^{-1}A^2P = P^{-1}AP\,P^{-1}AP$$
$$= \Lambda^2$$

and likewise any power of A. It will also diagonalize polynomials in A, since

$$P^{-1}\{aA^m + bA^n\}P = aP^{-1}A^mP + bP^{-1}A^nP$$
$$= a\Lambda^m + b\Lambda^n$$

and likewise convergent power series in A.

$$f(A) = a_0\mathbb{I} + a_1A + a_2A^2 + \cdots$$
$$P^{-1}f(A)P = P^{-1}\{A_0\mathbb{I} + a_1A + \cdots\}P$$
$$= a_0\mathbb{I} + a_1\Lambda + a_2\Lambda^2 + \cdots$$

$$= \begin{bmatrix} a_0+\lambda_1a_1+\lambda_1^2a_2+\cdots & 0 & \cdots & 0 \\ 0 & a_0+\lambda_1a_1+\lambda_2^2a_2+\cdots & & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & a_0+\lambda_na_1+\lambda_n^2a_2+\cdots \end{bmatrix}$$

$$= \begin{bmatrix} f(\lambda_1) & & \\ & f(\lambda_2) & \\ & & \ddots \; f(\lambda_n) \end{bmatrix}$$

$$= f(\lambda_1)\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & 0 \end{bmatrix} + f(\lambda_2)\begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & D \end{bmatrix}$$

$$+ \cdots + f(\lambda_n)\begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \cdots & & & \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

$$P^{-1}AP = f(\lambda_1) \,|1\rangle\langle 1| + f(\lambda_2)\,|2\rangle\langle 2| + \cdots + f(\lambda_n)\,|n\rangle\langle n|$$

$$P^{-1}AP = \sum_{i=1}^{N} f(\lambda_i)\,|i\rangle\langle i|$$

$$A = \sum_{i=1}^{N} f(\lambda_i)\; P\,|i\rangle\langle i|\,P^{-1}$$

But
$$P\,|i\rangle = [\,|1\rangle, |2\rangle \cdots |n\rangle\,] \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \end{bmatrix} = |i\rangle$$

And
$$\langle i|\,P^{-1} = [\,0 \;\; 0 \cdots 1 \cdots 0\,] \begin{bmatrix} \langle 1| \\ \langle 2| \\ \cdots \\ \langle n| \end{bmatrix} = \langle i|$$

$$\therefore \quad A = \sum_{j=1}^{N} f(\lambda_j)\,|j\rangle\langle j|$$

q.e.d.

Alternate Proof

Just substitute $A = \sum \lambda_i E_i$ in the power series for f and reduce.

Although stated under fairly restrictive conditions, Sylvester's theorem appears to be of more general validity. For instance it gives correctly negative and fractional powers, logarithms, and even functions which have been expanded in power series beyond their radius of convergence. The question which arises in such cases, however, is whether the results are unique.

No explicit process has yet been exhibited for calculating the eigenvectors of a matrix. Such a formula is possible, and can be derived by the aid of Sylvester's theorem.

Suppose that $P(x)$ is the characteristic polynomial of the matrix A. By definition, the eigenvalues satisfy

$$P(\lambda_i) = 0$$

so that if A has a complete set of eigenvectors

$$P(A) = \sum P(\lambda_i) \, |i\rangle\langle i|$$

will vanish identically. This is a special case of the Cayley-Hamilton theorem which states that every square matrix satisfies its own characteristic equation.

Being a polynomial, P(A) can be written in factored form:

$$P(A) = \prod_j \{ A - \lambda_j I \}$$

Let Γ_j be the following polynomial:

$$\Gamma_j = \prod_{i \neq j} \{ A - \lambda_i I \}$$

so that

$$P(A) = \Gamma_j (A - \lambda_j I) = (A - \lambda_j I) \Gamma_j$$

or, since $P(A) = 0$

$$A\Gamma_j = \lambda_j \Gamma_j$$

which resembles the equation for eigenvectors, except that Γ_j is a matrix, not a vector. It may be regarded as a row of columns, however, and the columns must be multiples of the j^{th} eigenvector, since

$$A\Gamma_j = A[\rho_1, \rho_2, ---, \rho_n]$$
$$= [A\rho_1, A\rho_2, ---, A\rho_n]$$

also

$$\lambda_j \Gamma_j = [\lambda_j \rho_1, \lambda_j \rho_2, -- \lambda_j \rho_n]$$

$$\therefore A\rho_1 = \lambda_j \rho_1$$
$$A\rho_2 = \lambda_j \rho_2$$
$$- -$$
$$A\rho_n = \lambda_j \rho_n$$

Using the other member of the equation above,

$$\Gamma_j A = \lambda_j \Gamma_j$$

It appears that the rows of Γ_j must be multiples of left eigenvectors:

$$T_j' A = \begin{bmatrix} \bar\sigma_1 \\ \bar\sigma_2 \\ -- \\ \bar\sigma_n \end{bmatrix} A = \begin{bmatrix} \bar\sigma_2 A \\ \bar\sigma_2 A \\ -- \\ \bar\sigma_n A \end{bmatrix}$$

$$\lambda_j T_j' = \begin{bmatrix} \lambda_j \bar\sigma_1 \\ \lambda_j \bar\sigma_2 \\ -- \\ \lambda_j \bar\sigma_n \end{bmatrix}$$

so that

$$\bar\sigma_1 A = \lambda_j \bar\sigma_1$$
$$\bar\sigma_2 A = \lambda_j \bar\sigma_2$$
$$--$$
$$\bar\sigma_n A = \lambda_j \bar\sigma_n$$

and if the eigenvectors are taken normalized as members of a biorthonormal set it appears that both the rows and columns of T_j are proportional to them, so that T_j itself must be proportional to an outer product of left eigenvectors and right eigenvectors. To discover the value of the proportionality constant, diagonalize T_j.

$$P^{-1} T_j P = P^{-1}\{\prod_{j \neq i} (A - \lambda_i \mathbb{I})\} P$$
$$= \prod_{i \neq j} P^{-1}\{A - \lambda_i \mathbb{I}\} P$$
$$= \prod_{i \neq j} \{\Lambda - \lambda_i \mathbb{I}\}$$
$$= \begin{bmatrix} \prod_{i \neq j} (\lambda_1 - \lambda_i) & 0 & -- & 0 \\ 0 & \prod_{i \neq j} (\lambda_2 - \lambda_i) & -- & 0 \\ 0 & 0 & -- & \prod_{i \neq j} (\lambda_n - \lambda_i) \end{bmatrix}$$

now, if $\lambda_i \neq \lambda_j$

$$\prod_{i \neq j} (\lambda_k - \lambda_i) = \delta_{jk} \prod_{i \neq j} (\lambda_j - \lambda_i)$$

since if $k \neq j$, λ_i will run through λ_k in the product,

$$P^{-1} T_j P = \| \delta_{\ell m} \delta_{mj} \prod_{i \neq j} (\lambda_j - \lambda_i) \|$$
$$= \prod_{i \neq j} (\lambda_j - \lambda_i) \; | j \rangle \langle j |$$

$$T_j = \prod_{i \neq j} (\lambda_j - \lambda_i) \, | j \rangle \langle j |$$

whereupon $$\| j \rangle\langle j \| = \frac{T_j}{\prod_{i \neq j} (\lambda_j - \lambda_i)}$$

$$| j \rangle\langle j | = \frac{\prod_{i \neq j} (A - \lambda_i I)}{\prod_{i \neq j} (\lambda_j - \lambda_i)}$$

this last matrix is the quantity which appears in Sylvester's theorem, so that it may be taken as it stands, or it may be factored to actually obtain eigenvectors.

The quantity $\tilde{X} A X = \sum_i \sum_j x_i a_{ij} x_j$ is called a quadratic form. $\tilde{X}^* A X$ is called a hermitean form, and so on. They are all scalars, and the hermitean form of a hermitean matrix will be real, since each term in the expansion appears with its complex conjugate. Real forms may be further classified according to their signs. Thus if for all vectors X, $\tilde{X}^* A X$ is positive, the form is called positive definite. Other classifications are

	negative definite	form < 0	for each $X \neq 0$
	positive semidefinite	form ≥ 0	" " " " "
and	positive definite	form > 0	" " " " "

Thus if A can be factored into the form $\tilde{B}^* B$, its hermitean form is positive semidefinite.

An important extremal property for bihermitean forms may be deduced: the quantity $\langle x | A | y \rangle$ assumes an extremal, for fixed $\langle x | y \rangle$, when is a left eigenvector of A and $| y \rangle$ is a right eigenvector of A.

To show this, consider $$\frac{\langle x | A | y \rangle}{\langle x | y \rangle}$$

form $$\frac{\langle x + \partial x | A | y + \delta y \rangle}{\langle x + \partial x | y + \partial y \rangle}$$, set the first order terms in δx, δy

equal to zero. This gives

$$\frac{\langle x|A|y\rangle + \langle\delta x|A|y\rangle + \langle x|A|\delta y\rangle + \langle\delta x|A|\delta y\rangle}{\langle x|y\rangle + \langle x|\delta y\rangle + \langle\delta x|y\rangle + \langle\delta x|\delta x\rangle}$$

$$= \left\{\frac{\langle x|A|y\rangle}{\langle x|y\rangle} + \frac{\langle\delta x|A|y\rangle}{\langle x|y\rangle} + \frac{\langle x|A|\delta y\rangle}{\langle x|y\rangle} + \frac{\langle\delta x|A|\delta y\rangle}{\langle x|y\rangle}\right\}\left\{1 - \frac{\langle x|\delta y\rangle}{\langle x|y\rangle} + \frac{\langle\delta x|y\rangle}{\langle x|y\rangle} + \alpha(x^2)\right\}$$

collecting first order differentials, equating them to zero:

$$-\frac{\langle x|A|y\rangle}{\langle x|y\rangle}\left\{\frac{\langle x|\delta y\rangle}{\langle x|y\rangle} + \frac{\langle\delta x|y\rangle}{\langle x|y\rangle}\right\} + \frac{\langle\delta x|A|y\rangle}{\langle x|y\rangle} + \frac{\langle x|A|\delta y\rangle}{\langle x|y\rangle} = 0$$

divide out $\langle x|y\rangle \neq 0$

$$\langle\delta x|\left[A - \frac{\langle x|A|y\rangle}{\langle x|y\rangle}\mathrm{I}\right]|y\rangle + \langle x|\left[A - \frac{\langle x|A|y\rangle}{\langle x|y\rangle}\mathrm{I}\right]|\delta y\rangle = 0$$

since $\langle\delta x|$ and $|\delta y|$ are arbitrary, both

$$\langle x|A = \langle x|\frac{\langle x|A|y\rangle}{\langle x|y\rangle}$$

$$A|y\rangle = \frac{\langle x|A|y\rangle}{\langle x|y\rangle}|y\rangle$$

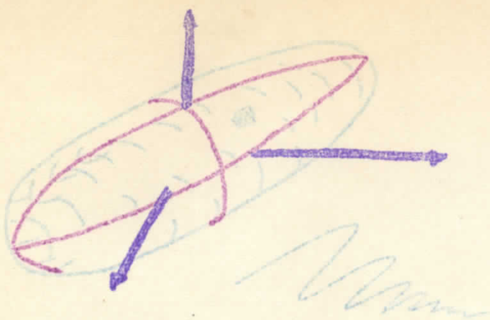
and $\langle x|$ and $|y\rangle$ must be eigenvectors with the eigenvalue

$$\frac{\langle x|A|y\rangle}{\langle x|y\rangle}.$$

The equation

$$\frac{\langle x|A|x\rangle}{\langle x|x\rangle} = \frac{1}{\langle x|x\rangle}$$

represents a quadratic surface in n dimensions, when A is hermitean. It is a skew quadratic since cross terms in the variables appear, but it is symmetric with respect to the origin. The eigenvectors point along the principal



axes, since they are the extremals of the quantity $\frac{\langle x|A|x\rangle}{\langle x|x\rangle}$ and therefore by the above equation of $\langle x|x\rangle$, the square of the distance to a point on the surface of the quadratic.

The eigenvalues of A are the reciprocals of the squares of the lengths of the semiaxes. To see this let the matrix be diagonalized by a matrix P, and substitute $X = PY$, so that

$$\langle x|A|x\rangle = \langle y|P^* A P|y\rangle = \langle y|\Delta|y\rangle$$

so that $\langle x|A|x\rangle = 1$ becomes

$$\lambda_1 |y_1|^2 + \lambda_2 |y_2|^2 + \cdots + |\lambda_n| y_n|^2 = 1$$

which is the normal form for the equation of a quadratic surface, while the eigenvalues are the reciprocals of the squares of the length of the semiaxes.

That a hermitean matrix always has an orthonormal system of eigenvectors can be deduced by considering its hermitean form. These vectors point along the semiaxes, and hence always exist and can be chosen orthogonal except in the case that the surface is a surface of revolution, and two or more semiaxes are equal. But then there is at least a plane of vectors, all of which make the form an extremal; and since they all belong to the same eigenvalue, the degeneracy already encountered with a multiple eigenvalue is reflected in this symmetry. However, in these planes one can always select two orthogonal vectors, even though not uniquely, and similarly with higher degeneracy.

From the diagonalized form of the hermitean form the nature of the surface may be deduced: a positive eigenvalue gives an elliptical section, zero eigenvalue a parabolic section, and a negative eigenvalue a hyperbolic one. Thus if A is a real hermitean matrix, $|x\rangle$ a real vector,

$$\langle x|A|x\rangle = 1$$

would define an ellipsoid in three-space if the three eigenvalues of A were positive; a hyperbolic paraboloid by two zero and one negative eigenvalues, and so on. If a matrix is positive definite all its eigenvalues are positive and the surface is closed. If the eigenvalues are non-negative it is positive semidefinite, and may have a parabolic part in addition.

A complete orthonormal set of vectors forms a suitable basis for a cartesian coordinate system. Thus, the eigenvectors of a hermitean matrix have the property that when they are chosen for a coordinate system, their matrix is diagonal, and its hermitean form is referred to its principal axes.

Hence diagonalization, which consists in applying a similarity transformation to a matrix, is nothing other than a transformation to a new axis system, possibly oblique, in which the matrix is diagonal. When the matrix is real and hermitean, this may be done by a rotation, a process which will be considered in detail later.

By the spectral theorem,

$$M = \sum \lambda_i |i\rangle\langle i|$$

$$\Lambda = P^{-1} M P = \sum \lambda_i \, P^{-1} |i\rangle\langle i| P$$

$$= \sum \lambda_i \; :i\rangle\langle i:$$

so that matrix P transforms the right eigenvectors to a system in which they appear as coordinate vectors, $:i\rangle$, while P^{-1} transforms the left eigenvectors similarly. Since the left eigenvectors, transform by the inverse matrix, P^{-1}, to that which transforms the right eigenvectors, the motivation for the name "reciprocal system" is seen.

In a sense, P transforms the columns of M , while P^{-1}, multiplied from the other side, transforms the rows, as may be seen by writing

$$M = \sum_{ij} a_{ij} \; :i\rangle\langle j:$$

$$P^{-1} M P = \sum_{i,j} a_{ij} \; P^{-1} :i\rangle\langle j: P$$

$$= \sum_{i,j} a_{ij} \, |i\rangle\langle j|$$

so that two matrices are needed to transform a matrix, while but one is needed to transform a vector.

The existence proof for eigenvectors just given was purely intuitive; it may be formalized as follows: $\langle x|x\rangle = 1$ defines a closed, bounded region, namely the surface of a unit hypersphere, $\sum |x_i|^2 = 1$. $\langle x|A|x\rangle$ is a continuous function of these x_i, hence attains a largest value; the $\langle x|_1$ for which this is true then being the eigenvector with largest eigenvalue. But the region defined by $\langle x|x\rangle = 1$, $\langle x|x_1\rangle = 0$ satisfies the same hypotheses, giving $\langle x|_2$, the eigenvector the next largest eigenvalue. $\langle x|x\rangle = 1$, $\langle x|x\rangle_1 = 0$, $\langle x|x\rangle_2 = 0$ is used to obtain another eigenvector until a complete set has been found. The process may fail when there are two largest such vectors, in which case there are infinitely many linearly dependent upon these two, and one of them may be chosen arbitrarily and the process continued. It will then not be unique.

The reason that the proof fails if A is not hermitean is that $\langle y|A|x\rangle$ is made extremal, with $\langle y|$ and $|x\rangle$ as eigenvectors. But, $\langle \overline{y}|^* \neq |x\rangle$ so that $\langle y|x\rangle = 1$ does not define a bounded region. It will succeed for any matrix whose left eigenvectors are the conjugates transposed of the right eigenvectors, however.

With an understanding of the significance of multiple eigenvalues, the formula for eigenvectors already obtained can be extended. The expression

$$T_j = \prod_{i \neq j} (\lambda_j - \lambda_i) |j\rangle\langle j| = \prod_{i \neq j} (A - \lambda_i \mathbb{I})$$

will fail to define $|j\rangle\langle j|$ if $\lambda_l = \lambda_j$, $l \neq j$.

If λ_j is a double eigenvalue, define the quantity

$$\Gamma_{jl} = \prod_{i \neq j,l} (A - \lambda_i I)$$

diagonalizing it,

$$P^{-1} \Gamma_{jl} P = \begin{bmatrix} \prod_{i \neq l,j} (\lambda_1 - \lambda_l) & & & \\ & \prod_{i \neq l,j} (\lambda_2 - \lambda_l) & & \\ & & --- & \\ & & & \prod_{i \neq l,j} (\lambda_n - \lambda_i) \end{bmatrix}$$

now

$$\prod_{i \neq j,l} (\lambda_k - \lambda_i) = 0 \qquad k \neq l,j$$

$$= \prod_{i \neq j,l} (\lambda_j - \lambda_i) \qquad k = l,j$$

$$P^{-1} \Gamma_{jl} P = \prod_{i \neq j,l} (\lambda_l - \lambda_i) \begin{bmatrix} 0 & & & \\ & 1_{jj} & & \\ & & \cdots 1_{ll} & \\ & & & \cdots 0 \end{bmatrix}$$

$$\Gamma_{j,l} = \prod_{i \neq l,j} (\lambda_l - \lambda_i) \{ |l\rangle\langle l| + |j\rangle\langle j| \}$$

But the diagonalization was not unique, and it is preferable to define

$$G_j = |l\rangle\langle l| + |j\rangle\langle j|$$

since

$$|l\rangle' = \cos\varphi\, |l\rangle + \sin\varphi\, |j\rangle$$

$$|j\rangle' = -\sin\varphi\, |l\rangle + \cos\varphi\, |j\rangle$$

gives

$$G'_j = \{ \cos\varphi\, |l\rangle + \sin\varphi\, |j\rangle \} \{ \cos\varphi \langle l| + \sin\varphi \langle j| \} + \{ -\sin\varphi\, |l\rangle + \cos\varphi\, |j\rangle \} \{ -\sin\varphi \langle l| + \cos\varphi \langle j| \}$$

$$= |l\rangle\langle l| + |j\rangle\langle j|$$

$$= G_j$$

so that G_j does not depend upon the particular diagonalization. Furthermore, since $\lambda_j |j\rangle\langle j| + \lambda_\ell |\ell\rangle\langle \ell| = \lambda_j G_j$ when $\lambda_j = \lambda_\ell$, Sylvester's theorem may be stated

$$f(A) = \sum_\lambda f(\lambda_j)\, G_j$$

where the G_j are outer products or sums of several outer products when there is a multiple eigenvalue, since a similar argument to the one given will hold when the multiplicity is greater than 2.

The G_j form an important class of operators, known as projective operators, which are said to project vectors into certain spaces. They project vectors onto unit vectors in a certain space, which gives a scalar, by

$$\langle i | X$$

This projection, onto a unit vector, is then made the component of the corresponding reciprocal vector to $\langle i|$, namely $|i\rangle$, giving the product

$$|i\rangle\, (\langle i | X)$$
$$= |i\rangle\langle i| X$$

Often one desires to think of a vector and its corresponding reciprocal vector as being really the same vector, with coordinates in different systems, since in matrix problems they both belong to the same eigenvalue, and for other reasons. One then calls the regular components of a vector its contravariant components, and the components of the reciprocal vector its contravariant components. This may be done for eigenvectors; for other vectors, their components in a given system are the covariant components; while if they are expressed in terms of the reciprocal system, the components are their contravariant components.

Upon knowing that a hermitean matrix can always be diagonalized, the following theorem can be proved:

Two hermitean matrices commute if and only if they can be diagonalized simultaneously.

To show sufficiency, let them both be diagonalized by U :

$$U^{-1}AU = \Lambda_1$$
$$U^{-1}BU = \Lambda_2$$
$$\Lambda_1\Lambda_2 = \Lambda_2\Lambda_1$$

since diagonal matrices commute.

$$U^{-1}AUU^{-1}BU = U^{-1}BUU^{-1}AU$$
$$U^{-1}ABU = U^{-1}BAU$$
$$AB = BA$$

q.e.d.

to show necessity. Suppose AB = BA

Let $$U^{-1}AU = \Lambda$$

since a hermitean matrix may be diagonalized. Now, define C by

$$U^{-1}BU = C$$
$$U^{-1}\Lambda U^{-1}UCU^{-1} = UCU^{-1}U\Lambda U^{-1}$$

since BA = AB

$$\therefore \Lambda C = C\Lambda$$

$$\begin{bmatrix} \lambda_1 C_{21} & \lambda_1 C_{12} & \cdots \lambda_1 C_{1n} \\ \lambda_2 C_{21} & \lambda_2 C_{22} & \cdots \lambda_2 C_{2n} \\ \cdots & & \\ \lambda_n C_{n1} & \lambda_n C_{n2} & \lambda_n C_{1n} \end{bmatrix} = \begin{bmatrix} \lambda_1 C_{11} & \lambda_2 C_{21} & \cdots \lambda_n C_{1n} \\ \lambda_1 C_{21} & \lambda_2 C_{22} & \cdots \lambda_n C_{2n} \\ \cdots & & \\ \lambda_1 C_{n1} & \lambda_2 C_{n2} & \cdots \lambda_n C_{nn} \end{bmatrix}$$

$$\lambda_i C_{ij} = \lambda_j C_{ij}$$

for $\lambda_i \neq \lambda_j$ $\quad C_{ij} = 0$

C is hermitean, since the eigenvectors of A, satisfy $\langle i| = \overline{|i\rangle}^*$

thus $$U^{-1} = \bar{U}^* \quad \text{Unitary}$$

$$\therefore \overline{\bar{U}^* B U^*} = \bar{U}^* \bar{B}^* U$$
$$= \bar{U}^* B U \; ; \quad C = \bar{C}^*$$

It is furthermore of the shape

$$\begin{bmatrix} \mu_1 & & & & \\ & \mu_2 & & & \\ & & \ddots & & \\ & & & \boxed{\begin{matrix} c \cdots c & c \cdots \\ c \cdots c \cdots c \cdots \\ c \cdots c \cdots c \cdots \end{matrix}} & \\ & & & & \ddots \\ & & & & & \mu_n \end{bmatrix}$$

while $U^{-1}AU$ is

$$\begin{bmatrix} \lambda_1 & & & & \\ & \lambda_2 & \ddots & & \\ & & & \boxed{\begin{matrix} \lambda_k & & \\ & \lambda_k & \\ & & \lambda_k \end{matrix}} & \\ & & & & \ddots \\ & & & & & \lambda_n \end{bmatrix}$$

so that C is mostly diagonalized. The undiagonalized submatrix of C is still hermitean, and may be diagonalized by Q. Then the matrix

$$\mathcal{Z} = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & \boxed{Q} & \\ & & & & \ddots \\ & & & & & 1 \end{bmatrix}$$

satisfies $\mathcal{Z}^{-1} = \bar{\mathcal{Z}}^*$, and the product $\mathcal{Z}^{-1} C \mathcal{Z}$ will be hermitean, and diagonal, while $\mathcal{Z}^{-1} U^{-1} A U \mathcal{Z} = U^{-1} A U$, and the diagonalization of A will be unaffected. Thus $\mathcal{Z}U$ will diagonalize both A and B. If there were several blocks of C left undiagonalized by U, corresponding to different sets of multiple eigenvalues, this amendment process would be repeated several times, diagonalizing each of the remaining blocks in turn.

q.e.d.

The fact that a matrix $\mathcal{T}$ may be necessary is due to the arbitrary choice of eigenvectors to diagonalize A, because of the multiple eigenvalue. In a certain subspace the eigenvectors may be chosen arbitrarily, hence need not coincide with those of B in the same subspace. But this matter is easily corrected by selecting instead the subeigenvectors of B, which is done by the matrix $\mathcal{T}$.

The amount by which two matrices fail to commute is called their commutator, denoted by the symbol

$$AB - BA = [A,B]$$

which is called a commutator bracket.

It is possible to solve an equation for an unknown inside a commutator bracket. If A and H are known matrices and M is unknown, let

$$[M,H] = A$$

$$MH - HM = A$$

If H can be diagonalized, let U be the matrix such that $U^{-1}HU = \Lambda$

$$U^{-1}MUU^{-1}HU - U^{-1}HNN^{-1}MU = U^{-1}AU$$

$$[U^{-1}MU]_{ij} = \langle i|M|j\rangle$$

$$\therefore \sum_k \langle i|m|k\rangle\langle k|\delta_{kj}\lambda_k|j\rangle - \sum_k \langle i|\delta_{ik}\lambda_k|k\rangle\langle k|M|j\rangle - \langle i|A|j\rangle$$

$$\lambda_j \langle i|M|j\rangle - \lambda_i \langle i|M|j\rangle = \langle i|A|j\rangle$$

$$\langle i|M|j\rangle = \frac{\langle i|A|j\rangle}{\lambda_j - \lambda_i}$$

and the elements of M in a known coordinate system are known.

When H has a multiple eigenvalue, $\lambda_j = \lambda_i$, say, the elements of M may be chosen arbitrarily, and any choice will give $\langle i|A|j\rangle = 0$. Likewise the diagonal elements of A must vanish, or no solution will be possible. When they vanish, the diagonal elements of H may be arbitrary, corresponding to the fact that to any solution of the commutator equation may be added a matrix commuting with H to give another solution. The proof will fall through also when H cannot be diagonalized.

Note that the vanishment of terms mentioned in the last paragraph occurs in the eigensystem of H, not in the system in which the problem was originally stated, which is commonly called the laboratory system.

The following commutator identifies are readily verified:

$$[M, AB] = [M, A]B + A[M, B]$$

$$[M, A] = -[A, M]$$

$$[M, A^{-1}] = -A^{-1}[M, A]A^{-1}$$

$$[M, A+B] = [M, A] + [M, B]$$

$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0$$

$$[M, A^n] = \sum_{1}^{n} A^{n-j}[M, A]A^{j-1}$$

< n^{th} commutator →

$$[M, [M, [M, \cdots [M, H] \cdots]]] = \sum_{0}^{n} \frac{n!}{j!(n-j)!} (-1)^j M^j H M^{n-j}$$

These formulae all resemble the formulae for derivatives with $[M,]$ replacing "$\partial/\partial M$". The reason for this is that commutator expressions, like derivatives, involve differences.

It is possible to represent certain functions of a matrix in terms of commutator series: thus if

$$U^{-1}HU = \Lambda$$

and there is an M such that $U^{-1} = e^{i\varepsilon M}$

$$e^{i\varepsilon M} H e^{-i\varepsilon M} = \Lambda$$

expanding in powers of ε

$$\left\{\mathbb{I} + \varepsilon i M + \frac{(i\varepsilon M)^2}{2!} + \cdots\right\} H \left\{\mathbb{I} - i\varepsilon M + \frac{(-i\varepsilon M)^2}{2!} + \cdots\right\} = \Lambda$$

$$H + i\varepsilon\{MH - HM\} + \frac{(i\varepsilon)^2}{2!}\{M^2H - 2MHM + HM^2\} + \cdots = \Lambda$$

$$H + i\varepsilon[M,H] + \frac{(i\varepsilon)^2}{2!}[M,[M,H]] + \frac{(i\varepsilon)^3}{3!}[M,[M,[M,H]]] + \cdots = \Lambda$$

which expresses the transformed H as a commutator series in M. For sufficiently small ε one has

$$H \doteq \Lambda - i\varepsilon[M,H]$$
$$\doteq \Lambda - [\ln U, H]$$

Under a similarity transformation, the determinant of a matrix remains unchanged. First note

$$|P^{-1}P| = |\mathbb{I}| = 1$$
$$|P^{-1}P| = |P^{-1}||P|$$
$$\therefore |P^{-1}| = |P|^{-1}$$

then
$$|P^{-1}AD| = |P^{-1}||A||P| = |A|$$

which proves the assertion. This means that the characteristic equation is unchanged since

$$|A - \lambda\mathbb{I}| = |P^{-1}|\ |A - \lambda\mathbb{I}|\ |P|$$
$$= |P^{-1}AP - \lambda\mathbb{I}|$$

and not only are the eigenvalues of a matrix independent of the coordinate system in which it is expressed, but so are the coefficients of the characteristic polynomial. Thus

$$\mathcal{L}(A) = \sum_i [A]_{ii} = \sum_i \lambda_i$$

is invariant, as was seen in examining the characteristic polynomial previously.

The result may also be proven by direct calculation, since

$$\mathcal{L}(P^{-1}AP) = \sum_j \sum_k \sum_\ell [P^{-1}]_{jk} [A]_{k\ell} [P]_{\ell j}$$

$$= \sum_k \sum_\ell \sum_j [A]_{k\ell} [P]_{\ell j} [P^{-1}]_{jk}$$

$$= \sum_k \sum_\ell [A]_{k\ell} \, \delta_{\ell k}$$

$$= \sum_k [A]_{kk} = \mathcal{L}(A)$$

The trace of the sum of two matrices is the sum of their traces, since

$$\mathcal{L}(A+B) = \sum [A+B]_{kk} = \sum \{ [A]_{kk} + [B]_{kk} \}$$
$$= \mathcal{L}A + \mathcal{L}B$$

while the trace of the product of two matrices is the same whichever order in which they are multiplied.

$$\mathcal{L}(AB) = \sum_k \sum_\ell [A]_{k\ell} [B]_{\ell k}$$
$$\mathcal{L}(BA) = \sum_\ell \sum_k [B]_{\ell k} [A]_{k\ell}$$

Then, the trace of a commutator is zero

$$\mathcal{L}([A, B]) = 0$$

In contrast to the invariance of eigenvalues, one has

$$Ax = \lambda x$$
$$P^{-1}APP^{-1}x = \lambda P^{-1}x$$
$$(P^{-1}AP)(P^{-1}x) = \lambda (P^{-1}x)$$

the same matrix which transforms A then transforms the eigenvectors to new coordinates. The eigenvectors are then called "covariant" rather than invariant.

Also since

$$|M| = |P^{-1}MP|, \qquad |M| = \prod_{i=1}^{n} \lambda_i$$

ORTHOGONAL AND UNITARY MATRICES

Vectors are regarded as having not only direction but magnitude. Having found the conditions under which the direction of a vector is left invariant it is natural to inquire for the circumstances under which magnitude is left invariant. If one considers instead the square of the length of a vector, rather than its length

$$|X|^2 = |MX|^2$$
$$\bar{X}X = \overline{MX}\,MX$$
$$= X\bar{M}MX$$

multiplying this out

$$\begin{aligned} x_1x_1 & & x_1a_{11}x_1 + x_1a_{12}x_2 + \cdots + x_1a_{1n}x_n \\ +x_2x_2 & = & +x_2a_{21}x_1 + x_2a_{22}x_2 + \cdots + x_2a_{2n}x_n \\ + \cdots & & + \cdots \\ +x_ny_n & & +x_na_{n1}y_1 + x_na_{n2}x_2 + \cdots + x_na_{nn}y_n \end{aligned}$$

since this is true for each n-dimensional vector X, set

$$X = |i\rangle \qquad a_{ii} = 1$$
$$X = |i\rangle + |j\rangle \qquad a_{ij} = -a_{ji} \qquad i \neq j$$

thus $\bar{M}M = I + N \qquad [N]_{ij} = a_{ij}$

and N is antisymmetric. But $\bar{M}M$ is symmetric;

$$\overline{\bar{M}M} = \bar{M}M = \overline{I + N} = I - N$$
$$\therefore N = 0$$
$$\bar{M}M = I$$
$$M^{-1} = \bar{M}$$

such a matrix is called orthogonal.

If M = $\begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ --- \\ \bar{X}_n \end{bmatrix}$ $\qquad \bar{M} = [X_1, X_2, ---, X_n]$

from

$$M\bar{M} = I$$

$$\bar{X}_i X_j = \delta_{ij} \qquad |M| \neq 0$$

and the rows of M form a complete orthonormal set. From $\bar{M}M = I$ the same property is seen to be true of the columns. Thus M is of the form

$$\begin{bmatrix} \cos\alpha_1 & \cos\alpha_2 & --- & \cos\alpha_n \\ \cos\beta_1 & \cos\beta_2 & --- & \cos\beta_n \\ --- & --- & & --- \\ \cos\eta_1 & \cos\eta_2 & --- & \cos\eta_n \end{bmatrix}$$

with direction cosines for elements. Since $\bar{M}M = I$, $|M|^2 = 1$, $|M| = \pm 1$ The determinant being non-zero, the rows are further guaranteed to form a complete set.

To avoid the difficulties concurrent with null-vectors, the hermitean length $\langle x|x \rangle$ of a vector has been introduced. If it is perserved

$$\langle x|x\rangle = \langle Ux|Ux\rangle$$
$$= \langle x|\bar{U}^*U|x\rangle$$

and by an argument similar to that above,

$$\bar{U}^*U = I$$
$$\bar{U}^* = U^{-1}$$

Such a matrix U is called unitary. By the same reasoning as before its rows and columns form a complete orthonormal set in the hermitean sense. Also,

$$|U|^*|U| = 1$$

and the absolute value of the determinant of such a matrix is 1.

Two theorems are true:

Theorem X. The eigenvalues of a unitary matrix are of absolute value 1.

Proof:

$$UX = \lambda X$$
$$\bar{X}^* \bar{U}^* = \lambda^* \bar{X}^*$$
$$\bar{X}^* \bar{U}^* U X = \lambda \lambda^* \bar{X}^* X$$
$$\bar{X}^* X (1 - \lambda \lambda^*) = 0$$
$$\therefore \lambda \lambda^* = 1$$

q.e.d.

Theorem XI: The left eigenvectors of a unitary matrix are the conjugate transposes of the right eigenvectors.

Proof:

$$UX = \lambda X$$
$$\bar{X}^* \bar{U}^* = \lambda^* \bar{X}^*$$
$$\bar{X}^* \bar{U}^* U = \frac{1}{\lambda} \bar{X}^* U$$
$$\bar{X}^* U = \lambda \bar{X}^*$$

q.e.d.

In fact, a unitary or hermitean matrix is diagonalized by a unitary matrix when the lengths of the eigenvectors are chosen so that $\langle \bar{U}^* = | i \rangle$

Of course these lengths need not be so chosen.

Two more theorems are true:

Theorem XII: If H is hermitean, then $e^{iH} = U$ is unitary.

Proof:

$$\overline{e^{iH}}^* = \overline{I + iH + \frac{1}{2!}(iH)^2 + \cdots}^*$$
$$= \bar{I}^* + \overline{iH}^* + \frac{1}{2!}(\overline{iH})^{*2} + \cdots$$
$$= I - iH + \frac{1}{2!}(-iH)^2 + \cdots$$
$$= e^{-iH}$$
$$= (e^{iH})^{-1}$$

q.e.d.

Since $\langle \bar{i} |^* = | i \rangle$ for unitary matrices, by the argument already given, there can be found a complete orthonormal set of eigenvectors for a unitary matrix.

The theorems listed above hold also for orthogonal matrices with the appropriate changes in nomenclature:

Theorem XIII. The determinant of an orthogonal matrix is $\pm$ 1.

Theorem XIV. Either the eigenvalues of an orthogonal matrix are $\pm$ 1 or they have null eigenvectors.

Theorem XV. If A is antisymmetric, then e^A is orthogonal.

Also, the following is true for hermitean matrices.

Theorem XVI. If H is Hermitean, $\{\mathbb{I} + iH\}\{\mathbb{I} - iH\}^{-1}$ is unitary.

FINITE ROTATIONS

Two cases of real unitary matrices are of particular interest: those are the two and three dimensional rotations and reflections. Let us first consider the two dimensional case. A matrix A

$$A = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}$$

having real coefficients is unitary when

$$A\bar{A} = \mathbb{I}$$

$$\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}\begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\alpha^2 + \beta^2 = 1$$

$$\gamma^2 + \delta^2 = 1$$

$$\alpha\gamma + \beta\delta = 0$$

these relations suggest setting

$$\alpha = \cos\theta \qquad \beta = \sin\theta$$

$$\gamma = \cos\phi \qquad \delta = \sin\phi$$

while the third equation requires

$$\cos\theta\cos\phi + \sin\theta\sin\phi = 0$$

$$\cos(\theta - \phi) = 0$$

$$\theta = \varphi + (k + \tfrac{1}{2})\pi$$

This gives two distinct forms for A, depending upon whether k is even or odd, namely

$$B = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \qquad C = \begin{bmatrix} -\cos\theta & \sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

which may be distinguished in that $|B| = +1$, $C = -1$

Taking C first, its characteristic equation is

$$\begin{vmatrix} -\cos\theta - \lambda & \sin\theta \\ \sin\theta & \cos\theta - \lambda \end{vmatrix} = 0$$

$$\lambda^2 = 1$$

$$\lambda_1 = 1 \qquad \lambda_2 = -1$$

The eigenvectors are obtained from the formula,

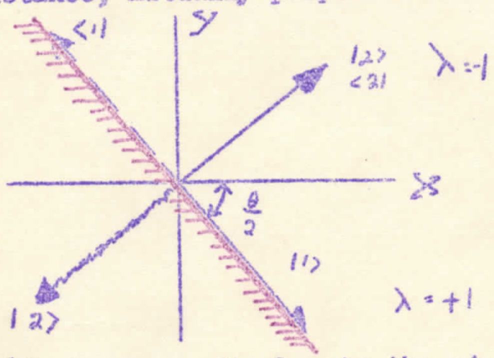
$$|1\rangle\langle 1| = \frac{1}{2}\begin{bmatrix} -\cos\theta - 1 & \sin\theta \\ \sin\theta & \cos\theta - 1 \end{bmatrix}$$

$$= \frac{1}{2}\begin{bmatrix} -2\cos^2\frac{\theta}{2} & 2\sin\frac{\theta}{2}\cos\frac{\theta}{2} \\ 2\sin\frac{\theta}{2}\cos\frac{\theta}{2} & -2\sin^2\frac{\theta}{2} \end{bmatrix}$$

$$= \begin{bmatrix} \cos\frac{\theta}{2} \\ -\sin\frac{\theta}{2} \end{bmatrix} \begin{bmatrix} -\cos\frac{\theta}{2} & \sin\frac{\theta}{2} \end{bmatrix}$$

while
$$|2\rangle\langle 2| = -\frac{1}{2}\begin{bmatrix} -\cos\theta + 1 & \sin\theta \\ \sin\theta & \cos\theta + 1 \end{bmatrix}$$
$$= \begin{bmatrix} \sin^2\frac{\theta}{2} & \sin\frac{\theta}{2}\cos\frac{\theta}{2} \\ \sin\frac{\theta}{2}\cos\frac{\theta}{2} & \cos^2\frac{\theta}{2} \end{bmatrix}$$
$$= \begin{bmatrix} \sin\frac{\theta}{2} \\ \cos\frac{\theta}{2} \end{bmatrix} \begin{bmatrix} \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix}$$

The eigenvectors bear out the conclusions of some of the theorems. They are, for instance, mutually perpendicular. Graphing them upon a plane, we have:



With the diagram the geometrical picture becomes clear. The transformation C makes a mirror reflection in a line tilted at an angle of $-\frac{\theta}{2}$ with respect to the X-axis.

Thus, $\langle 2|$ is perpendicular to the mirror, and if it originally points out of the mirror, $\langle 2|C$ points into the mirror. $\langle 1|$, lying parallel to the mirror is unchanged by the reflection.

The other type of orthogonal matrix, B, has eigenvalues given by
$$\begin{vmatrix} \cos\theta - \lambda & \sin\theta \\ -\sin\theta & \cos\theta - \lambda \end{vmatrix} = 0$$
$$\cos^2\theta - 2\lambda\cos\theta + \lambda^2 + \sin^2\theta = 0$$
$$\lambda = \frac{2\cos\theta \pm \sqrt{4\cos^2\theta - 4}}{2}$$
$$= \cos\theta \pm i\sin\theta$$
$$= e^{\pm i\theta}$$
$$\lambda_1 = e^{i\theta}$$
$$\lambda_2 = e^{-i\theta}$$

$$|1\rangle\langle 1| = \frac{1}{e^{i\theta} - e^{-i\theta}} \begin{bmatrix} \cos\theta - e^{i\theta} & \sin\theta \\ -\sin\theta & \cos\theta - e^{i\theta} \end{bmatrix}$$

$$= \frac{1}{2i\sin\theta} \begin{bmatrix} -i\sin\theta & \sin\theta \\ -\sin\theta & -i\sin\theta \end{bmatrix}$$

$$= -\frac{1}{2} \begin{bmatrix} 1 \\ -i \end{bmatrix} \begin{bmatrix} 1 & i \end{bmatrix}$$

$$|2\rangle\langle 2| = \frac{1}{e^{-i\theta} - e^{-i\theta}} \begin{bmatrix} \cos\theta - e^{-i\theta} & \sin\theta \\ -\sin\theta & \cos\theta - e^{-i\theta} \end{bmatrix}$$

$$= \frac{1}{-2i\sin\theta} \begin{bmatrix} i\sin\theta & \sin\theta \\ -\sin\theta & i\sin\theta \end{bmatrix}$$

$$= -\frac{1}{2} \begin{bmatrix} 1 \\ i \end{bmatrix} \begin{bmatrix} 1 & -i \end{bmatrix}$$

Since these eigenvectors do not have eigenvalues ± 1, they are null eigenvectors. Since they are complex, they cannot be drawn in a real diagram. But by observing the transformation

$$X' = BX$$

$$x' = x\cos\theta + y\sin\theta$$

$$y' = -x\sin\theta + y\cos\theta$$

one recognizes the familiar formula from analytic geometry for the rotation of a point in the plane by an angle θ.

The matric C was diagonalized by the matrix

$$\begin{bmatrix} \cos\frac{\theta}{2} & \sin\frac{\theta}{2} \\ -\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix} = M$$

which performs rotation by an angle $\frac{\theta}{2}$. Writing $C = M \Lambda M^{-1}$, with

$$\Lambda = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

we see that C operating on a vector is equivalent to rotating through an angle $\theta/2$, reflecting in the x-axis, and rotating the vector back again. A reflection can always be factored into a rotation and a reflection in the x-axis, either as above, or by writing

$$\begin{bmatrix} -\cos\theta & \sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

reflecting first in the y-axis and then rotating. Thus a two dimensional real unitary matrix represents either a rotation, or a combination of a rotation with a reflection, which is itself a single reflection, of course. Since the reflections have determinant -1, the rotations determinant +1, it is seen that two successive reflections, represented by the product of their matrices, gives a rotation; a rotation and a reflection, and two rotations, a rotation.

Note that since they have the same eigenvectors all rotations commute; but that reflections do not commute with each other or with rotations except in trivial cases, a conclusion which is confirmed by experience.

Sylvester's theorem may be invoked to find a matrix analogue to Euler's formula; $e^{ib} = \cos b + i \sin b$. Since

$$B = e^{i\theta} \; |1\rangle\langle 1| + e^{-i\theta} \; |2\rangle\langle 2|$$

write

$$\Gamma = \theta \; |1\rangle\langle 1| - \theta \; |2\rangle\langle 2|$$

so that

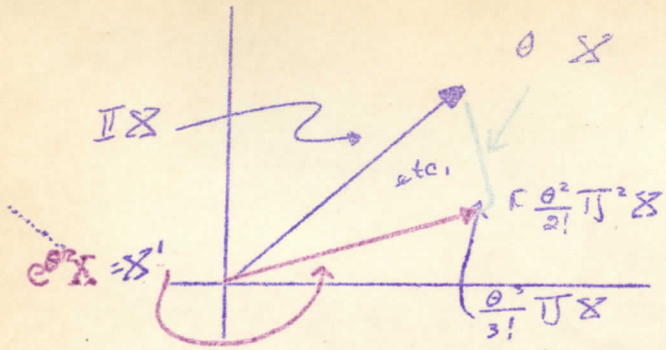
$$e^{i\Gamma} = B$$

but,

$$i\Gamma = \theta \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \theta \mathbb{J}$$

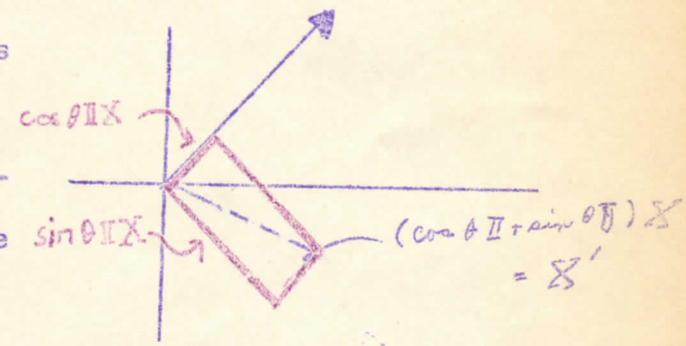
Now $\mathbb{J}^2 = -\mathbb{I}$ so that, expanding e^{in} by Taylor's series

$$\begin{aligned} e^{i\Gamma} &= \mathbb{I} + \mathbb{J}\,\theta + \mathbb{J}^2 \frac{\theta^2}{2!} + \cdots \\ &= \mathbb{I} + \mathbb{J}\,\theta - \frac{\theta^2}{2!}\mathbb{I} + \cdots \\ &= \mathbb{I}\left(1 - \frac{\theta^2}{2!} + \cdots\right) + \mathbb{J}\left(\theta - \frac{\theta^3}{3!} + \cdots\right) \\ &= \mathbb{I}\cos\theta + \mathbb{J}\sin\theta \end{aligned}$$



The diagram at the left illustrates the power series expansion as the sum of a series of operations on a vector X. JX is a vector perpendicular to X since $J\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ -x \end{bmatrix}$, so that for small angles $e^{\theta J} \doteq I + \theta J$ adds to a vector a small part of length $\theta |X|$ perpendicular to X. Since a tangent departs from a circle by terms of order θ^2, the length of X is left unchanged to first order by such an operation, and hence is a rotation to first order in θ.

The diagram at the right illustrates the matrix Euler's formula. The vector is now resolved into new components parallel and perpendicular to itself, and the new transformed vector formed with such coordinate axes. It is also readily interpreted by summing the perpendicular and parallel vectors of the other diagram separately.



The three dimensional case is similar. Recalling that the exponential of an antisymmetric matrix is orthogonal, write a 3 x 3 antisymmetric matrix.

$$\Gamma = \begin{bmatrix} 0 & a & -b \\ -a & 0 & c \\ b & -c & 0 \end{bmatrix}$$

Its characteristic equation is

$$\begin{bmatrix} -\lambda & a & -b \\ -a & -\lambda & c \\ b & -c & -\lambda \end{bmatrix} = -\lambda^3 + \lambda(a^2+b^2+c^2) = 0$$

whence

$$\lambda = 0, \pm i\sqrt{a^2+b^2+c^2}$$

making the substitution

$$\frac{a}{\sqrt{a^2+b^2+c^2}} = \cos\gamma \qquad \frac{b}{\sqrt{a^2+b^2+c^2}} = \cos\beta$$

$$\frac{c}{\sqrt{a^2+b^2+c^2}} = \cos\alpha$$

and

$$\sqrt{a^2+b^2+c^2} = \theta$$

$$\Gamma = \theta \begin{bmatrix} 0 & \cos\gamma & -\cos\beta \\ -\cos\gamma & 0 & \cos\alpha \\ \cos\beta & -\cos\alpha & 0 \end{bmatrix}$$

Its eigenvectors are obtained from

$$|1\rangle\langle 1| = \frac{(\Gamma - i\theta\mathrm{I})(\Gamma + i\theta\mathrm{I})}{(0-i\theta)(0+i\theta)} \qquad \lambda_1 = 0$$

$$= \begin{bmatrix} -i & \cos\gamma & -\cos\beta \\ -\cos\gamma & -i & \cos\alpha \\ \cos\beta & -\cos\alpha & -i \end{bmatrix} \begin{bmatrix} i & \cos\gamma & -\cos\beta \\ -\cos\gamma & i & \cos\alpha \\ \cos\beta & -\cos\alpha & i \end{bmatrix}$$

$$= \begin{bmatrix} 1-\cos^2\alpha-\cos^2\beta & \cos\beta\cos\alpha & \cos\gamma\cos\alpha \\ \cos\alpha\cos\beta & 1-\cos^2\alpha-\cos^2\gamma & \cos\beta\cos\gamma \\ \cos\alpha\cos\gamma & \cos\beta\cos\gamma & 1-\cos^2\beta\cos^2\alpha \end{bmatrix}$$

$$= \begin{bmatrix} \cos\alpha \\ \cos\beta \\ \cos\gamma \end{bmatrix} \begin{bmatrix} \cos\alpha & \cos\beta & \cos\gamma \end{bmatrix}$$

$$|2\rangle\langle 2| = \frac{\Gamma(\Gamma + i\theta\mathrm{I})}{(i\theta-0)(i\theta+i\theta)} \qquad \lambda_2 = i\theta$$

$$|3\rangle\langle 3| = \frac{\Gamma(\Gamma - i\theta\mathrm{I})}{(-i\theta-0)(-i\theta+i\theta)} \qquad \lambda_3 = -i\theta$$

Note: $|2\rangle\langle 2| = (|3\rangle\langle 3|)^*$. Their actual evaluation may be readily carried through, but they are not needed for the following discussion.

By Sylvester's theorem:

$$C = e^{T} = e^{0}\,|1\rangle\langle 1| + e^{i\theta}\,|2\rangle\langle 2| + e^{-i\theta}\,|3\rangle\langle 3|$$

since the last two terms are complex conjugates, C is real.

The eigenvector of C with eigenvalue 1 is real and gives the axis of rotation. Thus $(\cos\alpha, \cos\beta, \cos\gamma)$ are the direction cosines of the axis of rotation.

To obtain a geometrical interpretation of C it must be expanded by a formula analogous to Euler's. Calling

$$\mathcal{A} = \begin{bmatrix} 0 & \cos\gamma & -\cos\beta \\ -\cos\gamma & 0 & \cos\alpha \\ \cos\beta & -\cos\alpha & 0 \end{bmatrix}$$

$$\mathcal{A}^2 = \begin{bmatrix} 0 & \cos\gamma & -\cos\beta \\ -\cos\gamma & 0 & \cos\alpha \\ \cos\beta & -\cos\alpha & 0 \end{bmatrix}\begin{bmatrix} 0 & \cos\gamma & -\cos\beta \\ -\cos\gamma & 0 & \cos\alpha \\ \cos\beta & -\cos\alpha & 0 \end{bmatrix}$$

$$\mathcal{A}^2 = \begin{bmatrix} -\cos^2\gamma - \cos^2\beta & \cos\alpha\cos\beta & \cos\alpha\cos\gamma \\ \cos\beta\cos\alpha & -\cos^2\gamma - \cos^2\alpha & \cos\beta\cos\gamma \\ \cos\alpha\cos\gamma & \cos\gamma\cos\beta & -\cos^2\alpha - \cos^2\beta \end{bmatrix} = -\underline{\mathbb{I}} + |1\rangle\langle 1|$$

Since $\quad T = \theta\mathcal{A}$

$$e^{T} = e^{\theta\mathcal{A}} = \mathbb{I} + \theta\mathcal{A} + \frac{1}{2!}\theta^2\mathcal{A}^2 + \cdots$$

$$= \underline{\mathbb{I}} + \theta\mathcal{A} + \frac{\theta^2}{2!}(|1\rangle\langle 1| - \underline{\mathbb{I}}) + \frac{\theta^3}{3!}\mathcal{A}(|1\rangle\langle 1| - \underline{\mathbb{I}}) + \cdots$$

$$= \underline{\mathbb{I}}\left(1 - \frac{\theta^2}{2!} + \cdots\right) + \mathcal{A}\left(\theta - \frac{\theta^3}{3!} + \cdots\right) + |1\rangle\langle 1|\left(\frac{\theta^2}{2!} - \frac{\theta^4}{4!} + \cdots\right)$$

$$= \underline{\mathbb{I}}\cos\theta + \mathcal{A}\sin\theta + |1\rangle\langle 1|(1 - \cos\theta)$$

or $\quad C = \{\mathbb{I} - |1\rangle\langle 1|\}\cos\theta + \mathcal{A}\sin\theta + |1\rangle\langle 1|$

The quadratic form $\bar{X}\mathcal{A}Y$ will occur in the following discussion. It is

$$\sum_i\sum_j X_i \mathcal{A}_{ij} Y_j$$

But $\mathcal{A}_{ij} = (-1)^h |1\rangle_k$ $\quad i,j,k$ = some permutation of 1,2,3

where h is $\pm$ 1 depending upon whether i, j, k come in cyclical order or anticyclical order; and zero if two of 1,2,3 are repeated. Thus

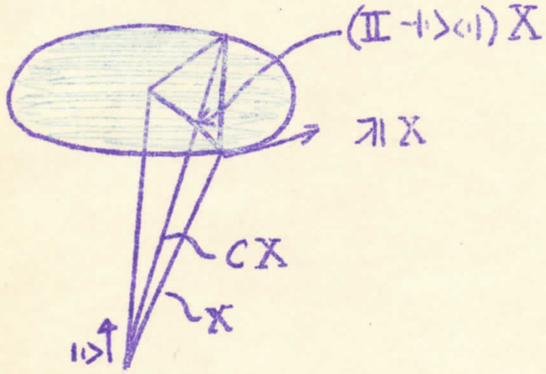
$$\bar{X}\mathcal{A}Y = \sum_{\phi} (-1)^h x_i y_j |1\rangle_k$$

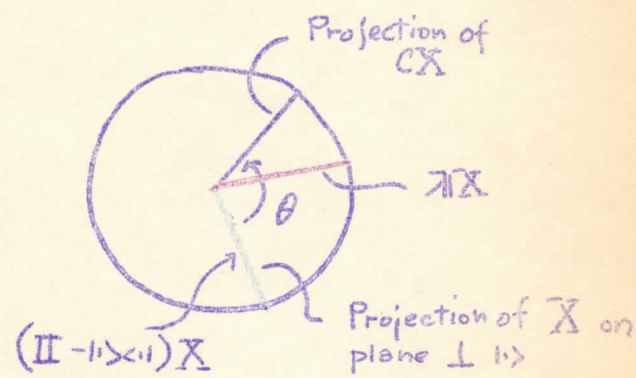
which is nothing other than the determinant

$$\begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ |2\rangle_1 & |2\rangle_2 & |2\rangle_3 \end{vmatrix} = \bar{X}\mathcal{A}Y$$

as may be veridied by expanding both sides of the equation if desired.

Thus $\mathcal{A}X$ is a vector perpendicular to $|1\rangle$ and X, since $\bar{X}\mathcal{A}X = 0$ as well as $\langle 1|\mathcal{A}X = 0$,





the product CX contains three terms:

$$\cos\theta\,\{\mathbb{I} - |1\rangle\langle 1|\}X$$

$$\sin\theta\,\mathcal{A}X$$

$$|1\rangle\langle 1|X$$

The latter is the portion of $\mathcal{X}$ parallel to the axis of rotation and is unchanged. The remaining two terms are written in terms of $\mathcal{N}\,\mathcal{X}$, perpendicular to $|1\rangle$ and $\mathcal{X}$ and hence in a plane perpendicular to $|1\rangle$, indicated by the red arrow in the diagram, $\{\mathbb{I} - |1\rangle\langle 1|\}\,\mathcal{X}$ is the part of $\mathcal{X}$ parallel to the plane perpendicular to $|1\rangle$, and is indicated by the green line. The previous discussion of the two dimensional case now applies.

That rotations occur in a plane may be further shown by partially diagonalizing C. Thus the matrix

$$[\,|1\rangle \quad \mathcal{X} \quad \mathcal{Y}\,] = S$$

where $\mathcal{X}$, $|1\rangle$ and $\mathcal{Y}$ are three mutually perpendicular unit vectors will transform C to a coordinate system in which the axis of rotation is the x axis. Notice that S itself is orthogonal, determinant + 1, and gives a rotation. However

$$\begin{bmatrix} \langle 1| \\ \bar{\mathcal{X}} \\ \bar{\mathcal{Y}} \end{bmatrix} [\,\mathbb{I}\cos\theta + \mathcal{N}\sin\theta + (1-\cos\theta)\,|1\rangle\langle 1|\,]\,[\,|1\rangle\ \mathcal{X}\ \mathcal{Y}\,]$$

$$= \begin{bmatrix} \langle 1| \\ \bar{\mathcal{X}} \\ \bar{\mathcal{Y}} \end{bmatrix} [\,|1\rangle \quad \mathcal{X}\cos\theta + \mathcal{N}\,\mathcal{X}\sin\theta \quad \mathcal{Y}\cos\theta + \mathcal{N}\,\mathcal{Y}\sin\theta\,]$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \bar{\mathcal{X}}\,\mathcal{N}\,\mathcal{Y}\sin\theta \\ 0 & \bar{\mathcal{Y}}\,\mathcal{N}\,\mathcal{X}\sin\theta & \cos\theta \end{bmatrix}$$

but if $|\,\mathcal{X}, \mathcal{Y}, |1\rangle$ form a right handed coordinate system, say, $\bar{\mathcal{X}}\,\mathcal{N}\,\mathcal{Y} = +1$

giving $\qquad \bar{\mathcal{Y}}\,\mathcal{N}\,\mathcal{X} = -1$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix}$$

which is a two dimensional rotation embedded in three-space.

This matrix is further diagonalized by

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & i & -i \end{bmatrix}$$

which contains a submatrix which diagonalizes the two dimensional rotation, so that

$$\begin{bmatrix} \cos\alpha & x_1 & y_1 \\ \cos\beta & x_2 & y_2 \\ \cos\gamma & x_3 & y_3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & i & -i \end{bmatrix} = \begin{bmatrix} \cos\alpha & x_1 + i y_1 & x_1 - i y_1 \\ \cos\beta & x_2 + i y_2 & x_2 - i y_2 \\ \cos\gamma & x_3 + i y_3 & x_3 - i y_3 \end{bmatrix}$$

diagonalizes C. This result is interesting in that it depends explicitly upon $\mathcal{X}$ and $\mathcal{Y}$, even though the eigenvectors of C are unique. We shall return to this point after completing the present discussion.

It now appears to be possible to give a prescription for producing a matrix which will rotate through a given angle θ about a given axis $(\cos\alpha, \cos\beta, \cos\gamma)$. If the following matrices are defined:

$$\Sigma_x = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}$$

$$\Sigma_y = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$\Sigma_z = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

then

$$C = e^{\theta\{\cos\alpha \Sigma_x + \cos\beta \Sigma_y + \cos\gamma \Sigma_z\}}$$

has the desired property. The converse problem is also soluble. The trace of the partially diagonalized matrix C was

$$\mathrm{Tr}\, \rho^{-1} C \rho = 1 + 2 \cos\theta$$

which by the trace theorem gives

$$\mathrm{Tr}\, \rho^{-1} C \rho = \mathrm{Tr}\, C = 1 + 2 \cos\theta$$

$$\theta = \left\{ \frac{\mathrm{Tr}\, C - 1}{2} \right\}$$

The eigenvectors giving the axis of rotation may be found by recalling.

$$C - \underline{I} \cos\theta + \lambda | \sin\theta + |\,\rangle\langle\,| (1 - \cos\theta)$$

which has the antisymmetric part $\mathcal{A}$, which already contains the direction cosines in question. Thus

$$\cos\alpha : \cos\beta : \cos\gamma : = c_{23} - c_{32} : c_{31} - c_{13} : c_{12} - c_{21}$$

Now, not only an antisymmetric matrix may be the logarithm of an orthogonal matrix, but certain other matrices as well. We had

$$\Gamma = \sum i\sigma_j |j\rangle\langle j|$$

where $e^{i\sigma_j}$ were the eigenvalues, necessarily of absolute value 1, of C. But it is also true that

$$\Gamma = \sum i(\sigma_j + 2\pi n_j) |j\rangle\langle j| \qquad n \text{ an integer}$$

satisfies

$$C = e^{\Gamma}$$

Now, the orthogonality condition was

$$C\tilde{C} = \mathbb{I}$$

$$e^{\Gamma}e^{\overline{\Gamma}} = \mathbb{I}$$

$$e^{\Gamma}e^{\tilde{\Gamma}} = \mathbb{I}$$

Now, a matrix commutes with its own inverse, hence these exponents must commute.

$$[\Gamma, \tilde{\Gamma}] = \emptyset$$

and

$$e^{\Gamma + \tilde{\Gamma}} = \mathbb{I}$$

Now, $\Gamma + \tilde{\Gamma}$ is twice the symmetric part of Γ , namely $2\Gamma^{(s)}$. This condition leaves the antisymmetric part completely arbitrary, which is a case already discussed.

$$\Gamma + \tilde{\Gamma} = \ln \mathbb{I}$$

$$2\Gamma^{(s)} = \ln\left(\sum_j e^{2\pi n_j i} |j\rangle\langle j|\right)$$

$$2\Gamma^{(s)} = \sum_j 2\pi n_j i |j\rangle\langle j|$$

Now, the $|j\rangle\langle j|$'s cannot be arbitrary, since

$$[\Gamma, \bar{\Gamma}] = 0$$

$$\Gamma = \Gamma^{(s)} + \Gamma^{(a)} \qquad \bar{\Gamma} = \Gamma^{(s)} - \Gamma^{(a)}$$

$$\Gamma\bar{\Gamma} = (\Gamma^s + \Gamma^a)(\Gamma^s - \Gamma^a)$$

$$\bar{\Gamma}\Gamma = (\Gamma^s - \Gamma^a)(\Gamma^s + \Gamma^a)$$

$$\Gamma\bar{\Gamma} = \bar{\Gamma}\Gamma \qquad \therefore \Gamma^s\Gamma^s - \Gamma^s\Gamma^a + \Gamma^a\Gamma^s - \Gamma^a\Gamma^a = \Gamma^s\Gamma^s + \Gamma^s\Gamma^a - \Gamma^a\Gamma^s - \Gamma^a\Gamma^a$$

$$\Gamma^a\Gamma^s = \Gamma^s\Gamma^a$$

and Γ^s and Γ^a commute, so that they must have a common set of eigenvectors. Unless $\theta = 0$, there is but one set for $\Gamma^{(a)}$, or $\Gamma^{(a)} = 0$, in this case the projective operators of $\Gamma^{(s)}$ are those of $\Gamma^{(a)}$ and when diagonalized, C must be

$$\begin{bmatrix} (-1)^{n_1} & 0 & 0 \\ 0 & (-1)^{n_2} & 0 \\ 0 & 0 & (-1)^{n_3} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & e^{i\theta} & 0 \\ 0 & 0 & e^{-i\theta} \end{bmatrix}$$

$$= \begin{bmatrix} (-1)^{n_1} & & \\ & e^{i(\theta + n_2\pi)} & \\ & & e^{i(-\theta + n_3\pi)} \end{bmatrix}$$

If only n_1 is odd, one has a reflection in the plane perpendicular to the axis of rotation. If n_2 and n_3 are both odd, there will be a rotation by 180° more than expected. If they are both even it makes no differende. The case where one is odd and one even cannot occur, since then one would have

$$C = e^0 |1\rangle\langle 1| + e^{i\theta} |2\rangle\langle 2| - e^{-i\theta} |3\rangle\langle 3|$$

which would give a non real matrix.

If $\theta = 0$, $e^{\Gamma e} = \mathbb{I}$, as with $\Gamma^a = 0$, $e^{0} = \mathbb{I}$. Thus the other case which may occur is a matrix with three eigenvalues ± 1.

The combination would be

	determinant	
1 1 1	+ 1	identity
1 1 -1	- 1	reflect xy plane
1 -1 1	- 1	" "
1 -1 -1	+ 1	180° rot x-axis
-1 1 1	- 1	reflect y-z plane
-1 1 -1	+ 1	180° rot y-axis
-1 -1 1	+1	180° rot z-axis
-1 -1 -1	- 1	180° rot, reflect y-z axis

so that as in the two dimensional case, matrices with determinant +1 are pure rotations; those with determinant -1 are mixtures of rotations and a reflection in the plane of rotation, or pure reflections in some plane.

SPINORS

It was pointed out that the matrix

$$\begin{bmatrix} \cos\alpha & x_1 + iy_1 & x_1 - iy_1 \\ \cos\beta & x_2 + iy_2 & x_2 - iy_2 \\ \cos\gamma & x_3 + iy_3 & x_3 - iy_3 \end{bmatrix}$$

diagonalized a pure rotational matrix, where $|1\rangle$ $\mathcal{Y}$ and $\mathcal{X}$ were three mutually perpendicular real unit vectors, $|1\rangle$ giving the axis of rotation.

Since $\mathcal{X}$ and $\mathcal{Y}$ may be chosen in a plane perpendicular to $|1\rangle$ with an arbitrary orientation, this freedom of choice seems to contradict the uniqueness of $|2\rangle$ and $|3\rangle$ as to direction. This is not the case as we now see. We have the condition that $|\bar{2}\rangle\, |2\rangle = 0$, since it is an eigenvector of an orthogonal matrix not belonging to $\lambda = \pm 1$. Writing $|2\rangle$ in terms of its real and imaginary parts

$$|2\rangle = W + iZ$$

$$\overline{|2\rangle}\,|2\rangle = \overline{(W+iZ)}\,(W+iZ)$$

$$= \overline{W}W - \overline{Z}Z + 2i\overline{W}Z = 0$$

equating real and imaginary parts,

$$\overline{W}W = \overline{Z}Z$$

$$\overline{W}Z = 0$$

so that W and Z are of the sa e length and mutually perpendicular, a condition true of Y and YZ above. The are perpendicular to $|1\rangle$, since

$$\langle 1|2\rangle = 0$$

$$\langle 1|2\rangle = \langle 1|W + i\langle 1|Z$$

and again equating real and imaginary parts to zero the desired result is obtained. $|2\rangle$ is, however, undetermined up to a phase factor, ie a scalar multiple $e^{-i\varphi}$, with real φ, since the eigenvectors may be multiplied by any scalar in virtue of their linearity. The only restriction we have imposed is that

$$\langle 2|2\rangle = 1$$

to obtain a biorthonormal set of eigenvectors. Now, since the matrix is orthogonal, and

$$\langle 2| = \overline{|2\rangle}^{*}$$

$$\overline{|2\rangle}^{*}|2\rangle = 1$$

a condition also satisfied by $e^{-i\varphi}|2\rangle$. But

$$e^{-i\varphi}|2\rangle = \{\cos\varphi - i\sin\varphi\}\{W + iZ\}$$

$$= \{W\cos\varphi + Z\sin\varphi\}$$

$$+ i\{-W\sin\varphi + Z\cos\varphi\}$$

so that W and Z have been rotated through an angle φ in the plane perpendicular to $|1\rangle$. This not only resolves the paradox but also permits another representation for finite rotations in three dimensions. Not only does each null vector provide an axis-system, namely W, Z, and $|1\rangle$, but transformations of the null vector have their effect upon the axis of the rotation, indicated by the vector $|1\rangle$.

Now, in fact, if 0 is an orthogonal matrix, $O|2\rangle = |2\rangle'$ transforms W and Z to a new axis system: $OW = W'$ $\quad OZ = Z'$, so that the same transformation must transform $|1\rangle$ $\quad O|1\rangle = |1\rangle'$

$|2\rangle$ is a null vector, so that its components are not independent. If

$$|2\rangle = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

then

$$a^2 + b^2 + c^2 = 0$$

calling

$$c = \pm\sqrt{(a - ib)(-a - ib)}$$

we have $\quad u = \sqrt{a - ib} \qquad v = \sqrt{-a - ib}$

$$\begin{cases} a = \dfrac{u^2 - v^2}{2} \\ b = i\,\dfrac{u^2 + v^2}{2} \\ c = -uv \end{cases}$$

where u and v are still undetermined with respect to a common $\pm$ sign. u and v have been deliberately chosen so that the following equation holds:

$$\begin{bmatrix} u \\ v \end{bmatrix} [u \ v] = \begin{bmatrix} a-ib & -c \\ -i & -a-ib \end{bmatrix} = P$$

Now, $|P| = -(a^2+b^2+c^2)$. The transformation $O|2\rangle$ in 3-space will cause the matrix P to vary. However, its determinant will remain constant, so that the transformation O on $|2\rangle$ may be replaced by a new transformation Q on P; $P' = Q^{-1}PQ$ which will transform the vectors $\begin{bmatrix} u \\ v \end{bmatrix}$, $[u, v]$, which are given the name "spinors". Thus a transformation Q on $\begin{bmatrix} u \\ v \end{bmatrix}$ corresponds to a transformation O on $|2\rangle$, and hence on $|1\rangle$.

Having discovered a new line of argument we now forsake the null vectors and consider transformations leaving the determinant $(x^2+y^2+z^2)$ invariant. Thus

$$|Q^{-1}PQ| = |P|$$

setting, unlike above

$$P = \begin{bmatrix} z & x-iy \\ x+iy & -z \end{bmatrix}$$

we have $\quad |P| = -(x^2+y^2+z^2)$

For the logarithm of Q write

$$-\mathfrak{Q} = \ln Q$$

$$Q^{-1}PQ = e^{\mathfrak{Q}} P e^{-\mathfrak{Q}}$$

and

$$P' = P + [\mathfrak{Q}, P] + \frac{1}{2!}[\mathfrak{Q}, [\mathfrak{Q}, P]] + \cdots$$

observe that P is of the form

$$P = x\sigma_x + y\sigma_y + z\sigma_z$$

where

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$$

$$\sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

with

$$\mathbb{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

These form a set of "unit vectors" which have the following multiplication table:

	$\mathbb{I}$	σ_x	σ_y	σ_z
$\mathbb{I}$	$\mathbb{I}$	σ_x	σ_y	σ_z
σ_x	σ_x	$\mathbb{I}$	$i\sigma_z$	$-i\sigma_y$
σ_y	σ_y	$-i\sigma_z$	$\mathbb{I}$	$i\sigma_x$
σ_z	σ_z	$i\sigma_y$	$-i\sigma_x$	$\mathbb{I}$

so that they obey the rules

$$\sigma_i^{\,2} = \mathbb{I}$$

$$\sigma_i \sigma_j = -\sigma_j \sigma_i$$

$$\sigma_i \sigma_j = i\sigma_k \qquad i\,j\,k \text{ in cyclical order}$$

so that

$$[\sigma_i, \sigma_j] = 2i\sigma_k \qquad \text{" \quad " \quad " \quad "}$$

with this we may proceed to establish a correspondence between the three dimensional rotations and the transformations of P. Writing

$$P' = \sum_i x_i \left\{ \sigma_i + [\mathcal{Z}, \sigma_i + \tfrac{1}{2!}[\mathcal{Z}, \mathcal{Z}, \sigma_i]] + \cdots \right.$$

and recalling the series

$$\mathcal{X}' = \sum_k \sum_i \frac{x_i \Gamma^k}{k!} : i \rangle$$

of the previous discussion. If the corresponding terms of the two series are compared, the first terms are

$$\sum_i x_i \sigma_i$$

$$\sum_i x_i : i \rangle$$

with the correspondence $\sigma_i \to : i \rangle$, which agrees with their respective choice, as unit vectors. The second terms are

$$\sum_i x_i [\mathfrak{z}, \sigma_i]$$

$$\sum_i x_i \Gamma : i \rangle$$

so that

$$\sum_j [\mathfrak{z}_j, \sigma_i] \to \Theta \{ \cos\alpha \Sigma_x + \cos\beta \Sigma_y + \cos\gamma \Sigma_z \}$$

If the right hand side is multiplied out, a multiplication table arises

$\bullet$	$:1\rangle$	$:2\rangle$	$:3\rangle$
Σ_x	$\mid 0\rangle$	$-:3\rangle$	$:2\rangle$
Σ_y	$:3\rangle$	$\mid 0\rangle$	$-:1\rangle$
Σ_z	$-:2\rangle$	$:1\rangle$	$\mid 0\rangle$

which must be matched with a commutator table for the $\mathfrak{z}'s$ and $\sigma's$.

$[\,,\,]$	σ_x	σ_y	σ_z
$\mathfrak{z}_1$	0	$-\sigma_z$	σ_y
$\mathfrak{z}_2$	σ_z	0	$-\sigma_x$
$\mathfrak{z}_3$	$-\sigma_y$	σ_x	0

Recalling the commutation rules for the σ's, it seems that this table is satisfied by

$$\mathcal{Z}_1 = -\frac{1}{2i}\sigma_x$$

$$\mathcal{Z}_2 = -\frac{1}{2i}\sigma_y$$

$$\mathcal{Z}_3 = -\frac{1}{2i}\sigma_z$$

The system is now closed, and the higher terms of the series will already agree by putting the σ's for the $\mathcal{Z}$'s. Thus

$$\mathcal{Z} = -\frac{\theta}{2i}\left\{\cos\alpha\,\sigma_x + \cos\beta\,\sigma_y + \cos\gamma\,\sigma_z\right\}$$

where the unit rotators are now also unit vectors.

Since $\sigma_i^2 = \mathbb{I}$, there is an exact analogue of Euler's formula in this representation. It might be pointed out in passing that this representation is called the "quaternion representation" since $\mathbb{I}, i\sigma_x, i\sigma_y, i\sigma_z$, multiply together as quaternions, in contrast to the other representation, called the vector representation.

Set $\qquad \sigma = \cos\alpha\,\sigma_x + \cos\beta\,\sigma_y + \cos\gamma\,\sigma_z$

then $\qquad \sigma^2 = \mathbb{I}$

and $\qquad e^{i\frac{\theta}{2}\sigma} = \mathbb{I}\cos\frac{\theta}{2} + i\sigma\sin\frac{\theta}{2}$

although

$$\begin{aligned} e^{-\mathcal{Z}} P e^{\mathcal{Z}} &= \left(\mathbb{I}\cos\frac{\theta}{2} + i\sigma\sin\frac{\theta}{2}\right) P \left(\mathbb{I}\cos\frac{\theta}{2} - i\sigma\sin\frac{\theta}{2}\right) \\ &= \cos^2\frac{\theta}{2} P + i\sin\frac{\theta}{2}\cos\frac{\theta}{2}\left\{\sigma P - P\sigma\right\} + \sin^2\frac{\theta}{2}\sigma P\sigma \\ &= \cos\theta\, P + \sin\theta\,\frac{[P,\sigma]}{2i} + (1-\cos\theta)\frac{P + \sigma P\sigma}{2} \end{aligned}$$

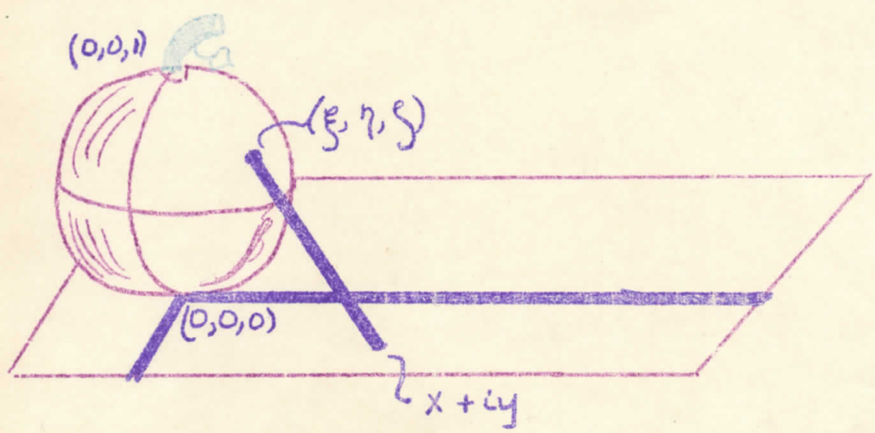
which may be compared with the expression obtained in the vector representation. To evaluate Q^{-1} explicitly, note $Q^{-1} = e^{\mathcal{Z}}$

$$\therefore\; Q^{-1} = \mathrm{I}\cos\frac{\theta}{2} + i\,\Phi\sin\frac{\theta}{2}$$

$$= \begin{bmatrix} \cos\frac{\theta}{2} + i\cos\alpha\sin\frac{\theta}{2}, & i\cos\alpha\sin\frac{\theta}{2} + \cos\beta\sin\frac{\theta}{2} \\ i\cos\alpha\sin\frac{\theta}{2} - \cos\beta\sin\frac{\theta}{2} & \cos\frac{\theta}{2} - i\cos\alpha\sin\frac{\theta}{2} \end{bmatrix}$$

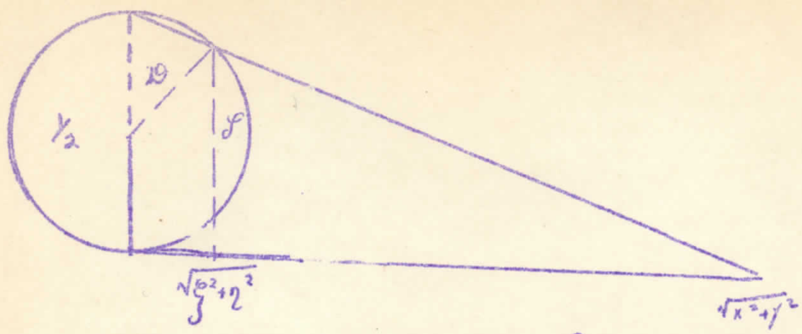
which is a unitary matrix.

There is a slight geometric connection which may be made between these two representations.



The polar stereographic projection maps the surface of a sphere of diameter 1 into the complex plane tangent at the south pole. Thus rotations, corresponding to transformations on the surface of the sphere are mapped into transformations in the plane. The quantities $(\xi, \eta, \zeta - \frac{1}{2})$ form the components of a vector; any transformation carrying one point of the sphere into another corresponds to an orthogonal transformation of the vector$(\xi, \eta, \zeta - \frac{1}{2})$. One may then investigate the corresponding transformation in the complex plane.

Calling ϑ the colatitude and drawing a section through the sphere:



$$\varphi = \arg(\xi + i\eta) \qquad \tfrac{1}{2}\sin\vartheta = \sqrt{\xi^2+\eta^2} = \tfrac{1}{2}\cos\vartheta + \tfrac{1}{2} = \zeta$$

$$\vartheta = \pi - 2\arctan\sqrt{x^2+y^2}$$

$$\zeta = \tfrac{1}{2}(\cos\vartheta + 1)$$

$$\xi = \tfrac{1}{2}\sin\vartheta\cos\varphi$$

$$\eta = \tfrac{1}{2}\sin\vartheta\sin\varphi$$

or, carrying out the requisite algebra

$$\zeta = \frac{x^2+y^2}{1+x^2+y^2}$$

$$\xi = \frac{x}{1+x^2+y^2}$$

$$\eta = \frac{y}{1+x^2+y^2}$$

since $\dfrac{\xi + i\eta}{\zeta} = \dfrac{x+iy}{x^2+y^2} = \dfrac{1}{z^*}$, we have

$$z = \frac{\zeta}{\xi - i\eta}$$

Now the variation of z can be found when $(\xi, \eta, \zeta - \tfrac{1}{2})$ is rotated. Recalling

$$C = \mathrm{I}\cos\theta + \mathcal{H}\sin\theta + (1-\cos\theta)\,|1\rangle\langle 1|$$

$$C = \begin{bmatrix} 2\sin^2\frac{\theta}{2}\cos^2\alpha + \cos\theta & \sin\theta\cos\gamma + 2\cos\alpha\cos\beta\sin^2\frac{\theta}{2} & 2\cos\alpha\cos\gamma\sin^2\frac{\theta}{2} - \cos\beta\sin\theta \\ 2\cos\alpha\cos\beta\sin^2\frac{\theta}{2} - \cos\gamma\sin\theta & 2\sin^2\frac{\theta}{2}\cos^2\beta + \cos\theta & 2\cos\beta\cos\gamma\sin^2\frac{\theta}{2} + \cos\alpha\sin\theta \\ 2\cos\alpha\cos\gamma\sin^2\frac{\theta}{2} + \cos\beta\sin\theta & 2\cos\beta\cos\gamma\sin^2\frac{\theta}{2} - \sin\theta\cos\alpha & 2\cos^2\gamma\sin^2\frac{\theta}{2} + \cos\theta \end{bmatrix}$$

Recalling $Q^{-1} = \begin{bmatrix} \cos\frac{\theta}{2} + i\cos\gamma\sin\frac{\theta}{2} & i\cos\alpha\sin\frac{\theta}{2} + \cos\beta\sin\frac{\theta}{2} \\ i\cos\alpha\sin\frac{\theta}{2} - \cos\beta\sin\frac{\theta}{2} & \cos\frac{\theta}{2} - i\cos\gamma\sin\frac{\theta}{2} \end{bmatrix}$

Calling
$$Q^{-1} = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}$$

we can show, omitting a large amount of algebra, that

$$C = \begin{bmatrix} \frac{1}{2}(\alpha^2 - \gamma^2 + \delta^2 - \beta^2) & \frac{i}{2}(\gamma^2 - \alpha^2 + \delta^2 - \beta^2 & \gamma\delta - \alpha\beta \\ \frac{i}{2}(\alpha^2 + \gamma^2 - \beta^2 - \delta^2) & \frac{1}{2}(\alpha^2 + \beta^2 + \gamma^2 + \delta^2) & -i(\alpha\beta + \gamma\delta \\ \beta\delta - \alpha\gamma & i(\alpha\gamma + \beta\delta) & \alpha\delta + \beta\gamma \end{bmatrix}$$

and with the aid of much more algebra, that if

$$z' = \frac{(\zeta - \frac{1}{2})' + \frac{1}{2}}{\xi' - i\eta'}$$

then
$$z' = \frac{\delta z - \gamma}{-\beta z + \alpha}$$

which gives the transformation of the complex plane corresponding to a rotation of the sphere. If one writes

$$z = \frac{\gamma_1}{\gamma_2}$$

the vector

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} = N$$

transforms by $\begin{bmatrix} \gamma_1' \\ \gamma_2' \end{bmatrix} = \begin{bmatrix} \delta & -\gamma \\ -\beta & \alpha \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}$

or, since Q^{-1} was unitary
$$\bar{N}' = \bar{N} Q$$

So to summarize, all the following transformations are concomitant

$$C\mathcal{X} = \mathcal{X}'$$
$$Q^{-1} P Q = P'$$
$$\frac{\delta z - \gamma}{-\beta z + \alpha} = z'$$
$$\bar{N} Q = \bar{N}'$$

in addition to the transformation by null vector, for which an explicit formula has not been given.

Rotations in higher dimensional spaces may also be discussed. Pure rotations are written as the exponentials of antisymmetric matrices. Now, for a real antisymmetric matrix, if

$$| A - \lambda \mathbb{I} | = 0$$
$$| \overline{A - \lambda \mathbb{I}} | = 0$$
$$| -A - \lambda \mathbb{I} | = 0$$

whence $| A + \lambda \mathbb{I}) = 0$, and where λ is an eigenvalue, so is $-\lambda$. Thus A may be diagonalized to

$$\begin{bmatrix} \lambda_1 & & & & \\ & -\lambda_1 & & & \\ & & \lambda_2 & & \\ & & & -\lambda_2 & \\ & & & & \ddots \end{bmatrix}$$

and can be written in terms of submatrices which will give two-dimensional rotations in a number of absolutely perpendicular subspaces. In an odd number of dimensions, there must always be one zero eigenvalue, and hence at least one real eigenvector for which $\lambda = 1$. Notice then that the rotations may be factored into a number of rotations about an $n-2$ dimensional axis, with all of the planes in which rotations occur absolutely perpendicular.

As in the two and three dimensional cases, orthogonal matrices with determinant $+1$ give pure rotations, those with determinant -1, combine reflections also.

A problem which may now be discussed is the following: A hypersphere, center fixed, is given two successive rotations about arbitrary axes and arbitrary angles. Is it so possible to choose the axes and angles that each point has been moved to a new position?

In two dimensions the problem is trivial. Take $\varphi_1 + \varphi_2 \neq 2\pi K$

In three dimensions, it is not soluble. For, if C_1 is one rotation and C_2 is another, $C_1 C_2$ is also a rotation, since

$$\overline{C_1 C_2}\, C_1 C_2 = \overline{C_2}\,\overline{C_1}\, C_1 C_2 = I$$

and

$$|C_1 C_2| = |C_1|\,|C_2| = +1$$

and in three dimensions there is an axis of rotation, so that no matter what is done, two points at least must remain fixed, the second rotation putting back two the first moved. However, in four dimensions the problem is again soluble. Take $C = e^{A}$ where

$$A = \begin{bmatrix} \lambda_1 & & & \\ & -\lambda_1 & & \\ & & \lambda_2 & \\ & & & -\lambda_2 \end{bmatrix}$$

then $C_1 = e^{A}$ $C_2 = e^{A_2}$, with

$$A_1 = \begin{bmatrix} 0 & & & \\ & 0 & & \\ & & \lambda_2 & \\ & & & -\lambda_2 \end{bmatrix} \qquad A_2 = \begin{bmatrix} \lambda_1 & & & \\ & -\lambda_1 & & \\ & & 0 & \\ & & & 0 \end{bmatrix}$$

then $C = e^{A_1 + A_2} = C_1 C_2$, since A_1, A_2 commute, but altogether $C = e^A$ has no real eigenvectors. Here the first rotation has been made about the $x-y$ plane, moving vectors in the $z-\omega$ plane. The second rotation now moves the vectors in the axis of the first, leaving unchanged the vectors in its own axis, which were all moved by the first.

If one considers plane rotations only the problem is no longer soluble for $n \geq 5$, since there are too many vectors sitting in various axes.

It may also be remarked in concluding that a hermitean matrix, having eigenvectors such that

$$\overline{|i\rangle}^* = \langle i|$$

is diagonalized by a unitary matrix, since

$$\overline{[\,|1\rangle\ |2\rangle\ \cdots\ |n\rangle\,]}^* = \begin{bmatrix} \langle 1| \\ \langle 2| \\ \cdots \\ \langle n| \end{bmatrix} = [\,|1\rangle, |2\rangle \cdots |n\rangle\,]^{-1}$$

so that a real hermitean matrix is diagonalized by a rotation to a suitable coordinate system, namely the orthonormal eigenvector system.