Mathematical Techniques
Part 4: Matrix Algebra

Lecture Notes by W.B. von Schlippe
formerly of Queen Mary and Westfield College, London, UK

Contents

1. Definition of matrix; matrix algebra. A matrix is an ordered two-dimensional array of numbers, the *matrix elements*, defined by the properties (I) to (IV). Before stating these properties, which define the matrix, we must introduce some notation and other definitions.

We shall usually denote matrices by capital letters A, B, …, Z, and the matrix elements by the corresponding lower case letters with two subscripts, such as a_{ij}, b_{ij} etc., where $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. Thus we have the following equivalent notations for a matrix:

$$A = (a_{ij}) = (a_{ij})_{mn} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \ldots & a_{3n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & a_{m3} & \ldots & a_{mn} \end{pmatrix} \tag{1}$$

This equation shows that the first subscript gives the row number; it is therefore called the *row index*. The second subscript is similarly the *column index*. The pair of numbers m and n defines the *order* or *dimension* of the matrix. If the matrix has m rows and n columns, it is called an $m \times n$ matrix (pronounced "m-by-n matrix").
If $n = m$ the matrix is called a *square matrix*; if $n \neq m$ it is called *rectangular matrix* except in the special cases $m = 1$, when it is called a *row matrix*, or $n = 1$, when it is called a *column matrix*. The matrix elements of row and column matrices are written with a single subscript.

Matrices are defined by the following properties:
Property I: Two $m \times n$ matrices $A = (a_{ij})$ and $B = (b_{ij})$ are equal iff $a_{ij} = b_{ij}$ for all $i \in [1, m]$ and $j \in [1, n]$.
Corollary: matrices of different order are not equal.
Property II: The sum of two $m \times n$ matrices $A = (a_{ij})$ and $B = (b_{ij})$ is the $m \times n$ matrix $C = (c_{ij})$ whose matrix elements are given by

$$c_{ij} = a_{ij} + b_{ij} \qquad \text{for all } i \in [1, m] \text{ and } j \in [1, n]$$

and is denoted by $C = A + B$.
Corollary: The sum of two matrices of different order is not defined.
Property III: The product of the $m \times n$ matrix $A = (a_{ij})$ and the number λ is the $m \times n$ matrix $C = (c_{ij})$ whose matrix elements are given by

$$c_{ij} = \lambda a_{ij} \qquad \text{for all } i \in [1, m] \text{ and } j \in [1, n]$$

and is denoted by $C = \lambda A$.
Corollary: the difference of two $m \times n$ matrices $A = (a_{ij})$ and $B = (b_{ij})$ is the sum of A and $(-1)B$ and is denoted by $C = A - B = A + (-1)B$.
Property IV: The product of $A = (a_{ij})_{mn}$ and $B = (b_{ij})_{pq}$ – in that order – is defined iff $p = n$; it is the matrix $C = (c_{ij})_{mq}$ whose matrix elements are given by

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \qquad \text{for all } i \in [1, m] \text{ and } j \in [1, q]$$

and is denoted by $C = AB$.
Corollary I: the matrix product AB is not defined if $p \neq n$.
Corollary II: the existence of the matrix product AB does not imply the existence of the product BA: the product BA is defined iff $q = m$. If both $p = n$ and $q = m$, and hence both products $C = AB$ and $D = BA$ are defined, then these two products will be equal only exceptionally.

Indeed, C and D will in general be of different order; they are of equal order only if A and B are square matrices. But even in the latter case C and D will in general be different. One says therefore that matrix multiplication is **noncommutative**. *In the exceptional case when $AB - BA = 0$ one says that A and B* commute.

Because of the importance of the order of the matrices in a matrix product, as expressed in Corollary II, one says that the matrix B is *premultiplied* by A in the product AB, and is *postmultiplied* by A in the product BA.[1]

Consider the matrices $A = (a_{ij})_{mn}$, $B = (b_{ij})_{np}$ and $C = (c_{ij})_{pq}$. Their orders are compatible in the sense that the matrix products AB and BC are defined. The product AB is an $m \times p$ matrix; it is therefore compatible with C for multiplication, i.e. the product $(AB)C$ exists. Similarly we can see that also the product $A(BC)$ exists. Important is the statement that the latter two products are equal; we can therefore drop the brackets and write

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

this means that matrix multiplication satisfies the *associative law*.
Matrix multiplication also satisfies the *distributive law*: if the sum of A and B and the product AC is defined, then the products $(A+B)C$ and BC are also defined and, moreover,

$$(A+B)C = AC + BC$$

The proof of the validity of the associative and distributive laws of matrix multiplication is left as an exercise for the reader.

2. Transpose, complex conjugate and Hermitian conjugate.

In addition to properties (I) to (IV), which define the algebra of matrices, we also define the operations of *transposition*, complex conjugation and *Hermitian conjugation* of matrices.

Transposition is the operation of exchanging the rows of a matrix with its columns. The matrix obtained by transposition from the $m \times n$ matrix $A = (a_{ij})_{mn}$ is called the *transpose* of A and denoted by A^T. A^T is an $n \times m$ matrix whose matrix elements are given by

$$A^T = (a_{ji})_{nm} = \begin{pmatrix} a_{11} & a_{21} & a_{31} & \dots & a_{m1} \\ a_{12} & a_{22} & a_{32} & \dots & a_{m2} \\ a_{13} & a_{23} & a_{33} & \dots & a_{m3} \\ \dots & \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & a_{3n} & \dots & a_{mn} \end{pmatrix}$$

Thus, for instance, the transpose of a 3×2 matrix is a 2×3 matrix:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}^T = \begin{pmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \end{pmatrix}$$

The **complex conjugate** of the matrix A is denoted by A^*; its matrix elements are the complex conjugates of the corresponding matrix elements of A.
Thus, for instance,

$$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}^* = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}$$

[1] some authors also say "the matrix A is multiplied on the right by B" referring to the matrix product AB or "is multiplied on the left by B" referring to BA.

Hermitian conjugation is the sequence of operations of complex conjugation and transposition (in either order). We shall denote the Hermitian conjugate of A by $A^\dagger$ (pronounced "A dagger"). Thus

$$A^\dagger = (A^*)^T = (A^T)^*$$

Important is the following **theorem:** the transpose of the matrix product AB is equal to the product $B^T A^T$.
Proof: we have $A^T = (a^T_{ij}) = (a_{ji})$ and similarly for B, hence, if $C = (c_{ij}) = AB$, then

$$c_{ij} = (AB)_{ij} = \sum_k a_{ik} b_{kj} = \sum_k a^T_{ki} b^T_{jk} = \sum_k b^T_{jk} a^T_{ki} = \tilde{c}_{ji}$$

where in the last step we have defined the matrix element $\tilde{c}_{ji}$ of a new matrix $\tilde{C}$. Thus we have

$$C = \tilde{C}^T$$

and hence $\tilde{C} = C^T$. But $C^T = (AB)^T$ and $\tilde{C} = B^T A^T$, hence

$$(AB)^T = B^T A^T$$

On the other hand, one can show, using again the definition of the matrix product, that the complex conjugate of AB is equal to the product of A^* and B^*, in that order. It follows that the Hermitian conjugate of AB is equal to the product of $B^\dagger$ and $A^\dagger$:

$$(AB)^\dagger = B^\dagger A^\dagger$$

3. Null matrix, diagonal matrix, unit matrix, triangular matrix; partitioned matrix, block diagonal and block triangular matrix.

There are many matrices which have special properties and therefore are given special names. Such special matrices will be introduced gradually. Here we introduce the following types of special matrices: the *null matrix*, *diagonal matrix*, *unit matrix*, *triangular matrix*, as well as the *partitioned matrix, block diagonal* and *block triangular matrix.*

3.1) A matrix is called *null matrix* if all its matrix elements are equal to zero.
We shall always denote null matrices simply by zero; it should be clear from the context what the order of a null matrix is.

An important property of the null matrix O is that, added to a matrix A of the same order, it leaves the matrix A unchanged:

$$A + O = A$$

3.2) A matrix is called *diagonal matrix* if its matrix elements are zero except some or all of its diagonal elements. Thus a diagonal matrix is of the form

$$A = \begin{pmatrix} a_{11} & 0 & 0 & \dots & 0 \\ 0 & a_{22} & 0 & \dots & 0 \\ 0 & 0 & a_{33} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & a_{nn} \end{pmatrix}$$

where not all of the diagonal elements a_{ii} are equal to zero. It is implicit in the definition of the diagonal matrix that it is a square matrix.

3.3) The diagonal matrix, whose diagonal elements are all equal to one, is called a *unit matrix*. Unit matrices can differ from each other in their orders. Thus, for instance, a 2×2 unit matrix is not equal to a 3×3 unit matrix, etc.
We shall usually denote unit matrices by E; it should always be clear from the context what the order of any particular unit matrix is. A convenient notation of the matrix elements of the unit matrix is in terms of the Kronecker-δ:

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

Thus we write, for instance, the 4×4 unit matrix as

$$E = (\delta_{ij})_{44} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

An important property of the unit matrix E is that, multiplied by a square matrix A of the same order, it leaves the matrix A unchanged:

$$EA = AE = A$$

Note that it is implicit in the latter equation that the $n \times n$ unit matrix E commutes with any $n \times n$ matrix A. This can be shown by using the definition of the matrix product:

$$(EA)_{ij} = \sum_{k=1}^{n} \delta_{ik} a_{kj} = a_{ij}$$

and similarly for the product AE.

3.4) Triangular matrix.
A matrix is called *upper triangular* if all matrix elements below the diagonal are equal to nought, but not all matrix elements above the diagonal are equal to nought.
A matrix is called *lower triangular* if all matrix elements above the diagonal are equal to nought, but not all matrix elements below the diagonal are equal to nought.

Examples of triangular matrices are

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{pmatrix} \qquad \text{upper triangular}$$

$$B = \begin{pmatrix} a_{11} & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \qquad \text{lower triangular}$$

3.5) Partitioned matrices, block diagonal and block triangular matrices.
Consider the following matrices, which play an important role in relativistic quantum mechanics:

$$\gamma_1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad \gamma_2 = \begin{pmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix}, \quad \gamma_3 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \qquad (2)$$

It is straight forward to carry out the basic operations of matrix algebra on this set of matrices. Thus, for instance,

$$a\gamma_1 + b\gamma_2 = \begin{pmatrix} 0 & 0 & 0 & a-ib \\ 0 & 0 & a+ib & 0 \\ 0 & a-ib & 0 & 0 \\ a+ib & 0 & 0 & 0 \end{pmatrix}$$

Multiplying the γ matrices is aslo straight forward but somewhat tedious. Having to work out the matrix product $\gamma_1\gamma_2\gamma_3$ is a fairly time consuming task. But this work is simplified if we take advantage of the structure of the γ matrices: the four matrix elements in the upper left-hand and in the lower right-hand corners form 2×2 null matrices; the four matrices in the upper right-hand and lower left-hand corners have simple symmetry properties. Let us denote the latter three 2×2 matrices by σ_1, σ_2 and σ_3, respectively, i.e.

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{3}$$

and write $\mathcal{O}$ for the 2×2 null matrix. Then we can write the γ matrices in the following compact form:

$$\gamma_1 = \begin{pmatrix} \mathcal{O} & \sigma_1 \\ \sigma_1 & \mathcal{O} \end{pmatrix}, \qquad \gamma_2 = \begin{pmatrix} \mathcal{O} & \sigma_2 \\ \sigma_2 & \mathcal{O} \end{pmatrix}, \qquad \gamma_3 = \begin{pmatrix} \mathcal{O} & \sigma_3 \\ \sigma_3 & \mathcal{O} \end{pmatrix}, \tag{4}$$

or even more concisely

$$\gamma_i = \begin{pmatrix} \mathcal{O} & \sigma_i \\ \sigma_i & \mathcal{O} \end{pmatrix}, \qquad i = 1,\, 2,\, 3 \tag{5}$$

The γ matrices in the form of Eqs. (4) or (5) are called *partitioned* matrices. The 2×2 matrices $\mathcal{O}$ and σ_i are called the *submatrices* of the γ matrices. The full beauty of partitioned matrices will be appreciated if we carry out all algebraic operations in the unpartitioned and in the partitioned form. Then we can see that the latter way is a great labour saving device. Consider the matrix product $\gamma_1\gamma_2$:

$$\gamma_1\gamma_2 = \begin{pmatrix} \mathcal{O} & \sigma_1 \\ \sigma_1 & \mathcal{O} \end{pmatrix} \begin{pmatrix} \mathcal{O} & \sigma_2 \\ \sigma_2 & \mathcal{O} \end{pmatrix} = \begin{pmatrix} \sigma_1\sigma_2 & \mathcal{O} \\ \mathcal{O} & \sigma_1\sigma_2 \end{pmatrix}$$

where we have applied the rule of matrix multiplication "row $\times$ column" to the submatrices. If we also realise that $\sigma_1\sigma_2 = i\sigma_3$, which is easily done by explicit multiplication, then we can write the result in the form of

$$\gamma_1\gamma_2 = \begin{pmatrix} i\sigma_3 & \mathcal{O} \\ \mathcal{O} & i\sigma_3 \end{pmatrix}$$

or, if we want to return to the unpartitioned form,

$$\gamma_1\gamma_2 = \begin{pmatrix} i & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \\ 0 & 0 & i & 0 \\ 0 & 0 & 0 & -i \end{pmatrix}$$

The partitioning shown in the above example is only a particular case. Sometimes it is convenient to partition a matrix by rows or by columns: see section 9. Other important cases are *block diagonal* and *block triangular* matrices. Thus consider the block diagonal matrix

$$M = \begin{pmatrix} a_{11} & a_{12} & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 & 0 \\ 0 & 0 & b_{11} & b_{12} & b_{13} \\ 0 & 0 & b_{21} & b_{22} & b_{23} \\ 0 & 0 & b_{31} & b_{32} & b_{33} \end{pmatrix}$$

which can be partitioned thus:

$$M = \begin{pmatrix} A & \mathcal{O} \\ \mathcal{O}^{\mathcal{T}} & B \end{pmatrix}$$

where

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \qquad B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}, \qquad \text{and} \quad \mathcal{O} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

This partitioning will be particularly useful if M is one of a set of matrices of similar structure on which one has to carry out algebraic manipulations.

An example of a block triangular matrix is the following:

$$N = \begin{pmatrix} a_{11} & a_{12} & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 & 0 \\ c_{11} & c_{12} & b_{11} & b_{12} & b_{13} \\ c_{21} & c_{22} & b_{21} & b_{22} & b_{23} \\ c_{31} & c_{32} & b_{31} & b_{32} & b_{33} \end{pmatrix}$$

which can be partitioned thus:

$$N = \begin{pmatrix} A & \mathcal{O} \\ C & B \end{pmatrix}$$

where

$$C = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \end{pmatrix}$$

and A, B and $\mathcal{O}$ have the previous meaning.

4. Inverse matrix.

We have so far defined three algebraic operations of matrices: addition, multiplication by numbers and matrix multiplication. We have also seen that the subtraction of matrices does not require a separate definition but can be reduced to addition and multiplication by -1. There is one algebraic operation still missing, namely division. Now, it turns out that there is no meaningful way of defining the division of two matrices. Instead one defines the nearest equivalent, namely the *inverse* of a matrix. Once we have the definition of the inverse it will become clear why one never speaks about the division of one matrix by another one.

The inverse of a matrix is defined only for *square* matrices. A rectangular matrix has no inverse.

Thus consider the $n \times n$ matrix A. Its inverse will be denoted by A^{-1} and defined by

$$AA^{-1} = E$$

where $E = (\delta_{ij})_{nn}$.

The inverse thus defined is called the *right inverse*, because we have multiplied A on the right by A^{-1}. We can also define the *left inverse* by

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

but we shall show that the left inverse is equal to the right inverse. It is therefore just called inverse, without the "left" or "right".
Proof: we have $AA^{-1} = E$ by definition, hence, if we premultiply by A^{-1}, we get $A^{-1}(AA^{-1}) = A^{-1}E = A^{-1}$, where in the last step we have used the fundamental property of the unit matrix. Now apply the associative law to the left-hand side, hence $A^{-1}(AA^{-1}) = (A^{-1}A)A^{-1}$, and then postmultiply by $A = (A^{-1})^{-1}$, hence

$$(A^{-1}A)[A^{-1}(A^{-1})^{-1}] = [A^{-1}(A^{-1})^{-1}]$$

Now, the expressions in square brackets are both equal to E by definition of the (right) inverse. We have therefore the result that

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

Note that our proof is still flawed: we have used the identity $A = (A^{-1})^{-1}$ without proof! To prove this statement we can proceed like this:
by definition $(A^{-1})^{-1}$ is the inverse of A^{-1}, therefore

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

and if we premultiply by A and use the associative law, then we get

$$A(A^{-1})(A^{-1})^{-1} = (AA^{-1})(A^{-1})^{-1} = AE = A$$

and since by definition $AA^{-1} = E$ we get the statement.

A most important property of the inverse matrix is its uniqueness. It must be emphasised that it is by no means obvious that the inverse matrix is uniquely defined. In fact, we know of inverse operations which are *not* uniquely defined, for instance integration, if considered as the inverse of differentiation. Thus it is necessary to prove the uniqueness of the inverse matrix.

We carry out the proof by making the converse assumption, namely that the inverse matrix is not uniquely defined, i.e. that we can have two matrices, A_1^{-1} and A_2^{-1} which are both the inverses of A, and then proceed to show that they are equal. Thus the assumption is that

$$AA_1^{-1} = E \qquad \text{and} \qquad AA_2^{-1} = E$$

and it follows that

$$AA_1^{-1} - AA_2^{-1} = 0$$

and hence, by the distributive law,

$$A(A_1^{-1} - A_2^{-1}) = 0$$

and if we premultiply this equation by A_1^{-1} (or by A_2^{-1}), then we get

$$A_1^{-1}A(A_1^{-1} - A_2^{-1}) = A_1^{-1} - A_2^{-1} = 0$$

Finally we also need a theorem concerning the inverse of the product of two square matrices:

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

Proof: By definition of the inverse we have

$$(AB)(AB)^{-1} = E$$

hence, if we premultiply by A^{-1} and apply the associative law, we get

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

but here the factor $(A^{-1}A)$ is equal to the unit matrix and can therefore be omitted. If we now premultiply by B^{-1} we get

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

and hence the statement.

Having defined the inverse matrix and established its most important properties we naturally want to find a method of actually calculating the inverse of a given matrix. We shall see that in order to do that we need yet another concept, namely the *determinant* of a matrix. Determinants arise by no means only in connection with the problem of finding the inverse of a matrix. So, if we turn in the next section to a discussion of determinants, it is with the understanding that they frequently arise in numerous applications.

5. Determinants.

5.0) Introduction.

5.1) Permutations.

5.2) Definition of the determinant of a square matrix.

5.3) Properties of determinants.

- 5.3.1) Determinant of the transpose.
- 5.3.2) Interchange of rows or columns.
- 5.3.3) Multiplication of a determinant by a number.
- 5.3.4) Determinant with two equal rows or columns.
- 5.3.5) Sum of determinants.

5.4) Determinant of diagonal, skew-diagonal and triangular matrices.

5.5) Determinant of block-diagonal, block-skew-diagonal and block-triangular matrices.

5.6) Product of determinants.

5.7) Minors and cofactors of a matrix; expansion theorem.

5.8) Inverse matrix.

5.9) A note on the practical evaluation of determinants.

5.10) Exercises.

5.0) Introduction.

The notion of a determinant applies only to square matrices. In the simplest case of a 2×2 matrix $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ the determinant is defined by

$$\det A = a_{11}a_{22} - a_{12}a_{21}$$

The definition of the determinant of a square matrix of arbitrary dimension will be given in section 5.2. Before formulating this definition we need another useful concept, that of *permutations.*

5.1) Permutations.

A permutation of a sequence of numbers is a sequence of the same numbers in a different order. Thus, the sequence $(3, 2, 1)$ is a permutation of the sequence $(1, 2, 3)$. To characterize a permutation completely it is necessary to specify both the initial and the final sequences. A convenient notation to do that is to write

$$\begin{pmatrix} initial & sequence \\ final & sequence \end{pmatrix}$$

i.e. in the previous example we write $\begin{pmatrix} 1,\ 2,\ 3 \\ 3,\ 2,\ 1 \end{pmatrix}$

Any permutation of a sequence of numbers is obtained from the initial sequence by a succession of interchanges of pairs of elements of the sequence. In the above example there is one interchange of the pair 3, 1. The succession of interchanges that lead to a given permutation of a sequence is not unique. In our example we can get the same permutation by first interchanging the pair 1, 2, giving the sequence (2, 1, 3), then in the new sequence interchanging the pair 1, 3 to give (2, 3, 1), and finally interchanging the pair 2, 3. Note that the first method required *one* interchange whereas the second method needed *three* interchanges.

A fundamental property of permutations is that a certain permutation $(k_1,\ k_2,\ \ldots,\ k_n)$ of the sequence $(1,\ 2,\ \ldots,\ n)$ can be obtained only either by an even number or by an odd number of interchanges. Correspondingly the permutation is called *even* or *odd.*

A particular case of a permutation is that which leaves the initial sequence unchanged. This requires zero permutations; it is therefore counted among the even permutations.

Example 1: The sequence (1, 2, 3) has six permutations, three of which are even and three are odd; they are

(1, 2, 3)	(2, 3, 1)	(3, 1, 2)	even
(1, 3, 2)	(3, 2, 1)	(2, 1, 3,)	odd

To verify the assignments even and odd to these permutations is left as an exercise for the reader.

It is generally true that half the permutations of any sequence are even and half are odd.

One can prove that the number of permutations of a sequence of N numbers is $N!$. The proof is accomplished by induction: for $N = 1$ the sequence is (1) and there is one permutation. For $N = 2$ we have the sequence (1, 2), which itself is an even permutation, and one additional permutation, namely (2, 1), which is an odd permutation; the number of permutations of two numbers is therefore $2 = 2!$. For $N = 3$ we proceed as follows: keep one of the numbers fixed; for the remaining two numbers there are two permutations; since we can keep each of the three numbers fixed, the total number of permutations is $2! \cdot 3 = 6 = 3!$. By the same reasoning we can show that the statement is true for any N if it is true for $N - 1$, which completes the proof.

To write down the definition in mathematical notation it will be convenient to introduce the symbol $\varepsilon_{k_1 k_2 \ldots k_n}$, which is defined to be $+1$ (-1) if $[k] \equiv (k_1 k_2 \ldots k_n)$ is an even (odd) permutation of the natural sequence $(1,\ 2,\ \ldots,\ n)$ and zero otherwise:

$$\varepsilon_{k_1 k_2 \ldots k_n} = \begin{cases} +1 & \text{if } [k] \text{ is even} \\ -1 & \text{if } [k] \text{ is odd} \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

Corollary: The ε symbol changes sign if two of its subscripts are interchanged.
Indeed, if the sequence $(k_1 k_2 \ldots k_{i-1} k_i k_{i+1} \ldots k_{j-1} k_j k_{j+1} \ldots k_n)$ is an even (odd) permutation of the natural sequence, then $(k_1 k_2 \ldots k_{i-1} k_j k_{i+1} \ldots k_{j-1} k_i k_{j+1} \ldots k_n)$ is an odd (even) permutation, and therefore the corresponding ε symbols have opposite signs.

We are now ready to formulate the definition of the determinant of a square matrix.

5.2) Definition of the determinant of a square matrix. The determinant of the square matrix A is denoted by $\det A$. Also used is the notation

$$\det A = \begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{vmatrix}$$

Definition: The determinant of the $n \times n$ matrix $A = (a_{ij})$ is the sum of all products of the form $a_{1k_1} a_{2k_2} \ldots a_{nk_n}$, where $(k_1,\ k_2,\ \ldots,\ k_n)$ are all permutations of the sequence $(1,\ 2,\ \ldots,\ n)$; terms for which $(k_1,\ k_2,\ \ldots,\ k_n)$ is an even (odd) permutation enter into the sum with a plus (minus).

Using the ε symbol we can restate the definition of the determinat of A in the following simpler form:

$$\det A = \sum_{[k]} \varepsilon_{k_1 k_2 \ldots k_n} a_{1k_1} a_{2k_2} \ldots a_{nk_n} \tag{7}$$

where the $\sum\limits_{[k]}$ means that the sum is taken over all permutations $k_1 k_2 \ldots k_n$.

It is very useful to remember that *each term of the sum* $\sum_{[k]} \varepsilon_{k_1 k_2 \ldots k_n} a_{1k_1} a_{2k_2} \ldots a_{nk_n}$ *has one and only one matrix element from every row and column of the matrix as a factor.*

Example 2: consider the 3×3 matrix $A = (a_{ij})_{nn}$. The determinant of A is

$$\det A = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{n1} & a_{n2} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33}$$

Note that the terms of $\det A$ in the example can be grouped thus:

$$\det A = a_{11}(a_{22}a_{33} - a_{23}a_{32}) + a_{12}(a_{23}a_{31} - a_{21}a_{33}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$$

and in the expressions in brackets we recognise determinants of 2×2 matrices. i.e.

$$\det A = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

i.e. each element a_{1j} of the top row appears multiplied by the determinant, obtained from $\det A$ by crossing out the row and column that contains a_{1j}, and enters with a sign given by the sequence $(+, -, +)$. This rule is referred to as expansion about the first row. The terms of the determinant can be grouped in different ways, for instance like this:

$$\det A = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21}\begin{vmatrix} a_{12} & a_{33} \\ a_{32} & a_{13} \end{vmatrix} + a_{31}\begin{vmatrix} a_{12} & a_{23} \\ a_{22} & a_{13} \end{vmatrix}$$

referred to as expansion about the first column. Similarly one can expand about any row or column of the determinant.

The expansion about rows (or columns) can be repeated with the determinants of lower order, that arise after the first expansion, and continued until, in the last step, all 2×2 determinants have been evaluated.

5.3) Properties of determinants.

5.3.1) Determinant of the transpose.
The determinant of the transpose A^T of A is equal to the determinant of A:

$$\det A^T = \det A \tag{8}$$

Indeed, if we expand $\det A$ about its rows and A^T about its columns, the two expressions coincide term for term.

5.3.2) Interchange of rows or columns.
A determinant changes sign if two adjacent rows (or columns) are interchanged, for instance:

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

This can be seen most easily from the definition of the ε symbol: exchange of two rows corresponds to the exchange of two indices of $\varepsilon_{k_1 k_2 \dots k_n}$.

5.3.3) Multiplication of a determinant by a number.

$$\lambda \det A = \det A' \tag{9}$$

where the matrix A' differs from A in that any one of its rows or columns is multiplied by λ.

Corollary: given the $n \times n$ matrix $A = (a_{ij})_{nn}$ and the number λ, we have

$$\det(\lambda A) = \lambda^n \det A$$

Example 3:

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

5.3.4) Determinant with two equal rows or columns.
The determinant of A is zero if two of its rows or columns are proportional to each other element by element.
We illustrate this property with an example. It is not difficult to convince oneself that the statement is generally true.

Example 4:

$$\begin{vmatrix} 1 & 2 & 3 & 5 \\ 2 & 3 & 1 & -1 \\ 3 & 6 & 9 & 15 \\ 1 & 3 & 2 & 7 \end{vmatrix} = 0$$

because the first and third rows are proportional.

Corollary 1: $\det A = 0$ if two rows or columns of A coincide.
Corollary 2: $\det A = 0$ if one row or column of A has only null elements.

5.3.5) Sum of determinants.
Consider the matrix $A = (a_{ij})$ and a second matrix A', which coincides with A in all matrix elements except those of one row (or column), e.g.

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}, \qquad A' = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ b_{i1} & b_{i2} & \dots & b_{in} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

then

$$\det A + \det A' = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{i1}+b_{i1} & a_{i2}+b_{i2} & \dots & a_{in}+b_{in} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

The proof follows directly from the definition of the determinant.

Corollary 3: the determinant of A is unchanged if one adds the multiple of one row (or column) to another row (or column).

The latter property is of utmost importance for the practical evaluation of determinants. In fact, to evaluate a determinant of an order higher than the second, one does not proceed from the definition (7) or by way of the expansion etc., but rather by applying a sequence of elementary operations to bring the determinant into a triangular form, which is equal to the product of the diagonal elements (see below, section 5.4). By elementary operation one means any one of the operations, discussed in this section, that leaves the determinant unchanged: exchange of rows or columns, addition of the multiple of one row or column to another row or column. A practical application of this strategy will be given below in Example 5.

5.4) Determinant of diagonal, skew-diagonal and triangular matrices.

The determinant of a diagonal matrix is equal to the product of its diagonal elements. Indeed, the product of the diagonal elements is the only nonzero term in the sum (7).

The determinant of a skew-diagonal matrix is equal in magnitude to the product of its skew-diagonal elements, with a positive (negative) sign if the are an odd (even) number of factors. The first part of the statement can be seen as in the previous case. The sign follows from the definition of the ε symbol: if S is a skew-diagonal $n \times n$ matrix, we have

$$\det S = \varepsilon_{n,n-1,\dots,1} a_{1n} a_{2,n-1} \dots a_{n1} \tag{10}$$

i.e. the sign is determined by the permutation

$$\begin{pmatrix} 1 & 2 & \dots & n \\ n & n-1 & \dots & 1 \end{pmatrix}$$

which is even (odd) if n is odd (even).

Determinant of a triangular matrix.
Recall the definition of triangular matrices given in Section 3.4:

> A matrix is called *upper triangular* if all matrix elements below the diagonal are equal to nought, but not all matrix elements above the diagonal are equal to nought.
> A matrix is called *lower triangular* if all matrix elements above the diagonal are equal to nought, but not all matrix elements below the diagonal are equal to nought.

The determinant of a triangular matrix is equal to the product of its diagonal elements. Indeed, recall the rule given in section 5.2: *each term of the sum* $\sum_{[k]} \varepsilon_{k_1 k_2 \dots k_n} a_{1k_1} a_{2k_2} \dots a_{nk_n}$ *has one and only one matrix element from every row and column of the matrix as a factor.* Now, if we choose a nonzero off-diagonal element, then that product of elements must necessarily also have a factor of nought. Therefore the only nonzero product of the sum is that which has only diagonal elements.

Alternatively we can consider a particular case. Let $A = (a_{ij})_{44}$ be an upper triangular matrix and evaluate its determinant by expanding about the first column:

$$\begin{aligned} \det A &= \begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} & a_{24} \\ 0 & a_{33} & a_{34} \\ 0 & 0 & a_{44} \end{vmatrix} \\ &= a_{11} a_{22} \begin{vmatrix} a_{33} & a_{34} \\ 0 & a_{44} \end{vmatrix} = a_{11} a_{22} a_{33} a_{44} \end{aligned}$$

where in each subsequent step we have continued expanding about the first column of the remaining determinant. The statement obviously generalises to triangular determinants of arbitrary order.

To prove the theorem also for lower triangular matrices we have to repeat the argument but expand in each step about the first row.

Example 5: to illustrate the properties of determinants we shall use them now to evaluate a 3×3 matrix in different ways.
Consider the matrix

$$A = \begin{pmatrix} 1 & 7 & 5 \\ -4 & 4 & 8 \\ 2 & 6 & 9 \end{pmatrix}$$

and let us first find its determinant by expansion about the top row, followed by the evaluation of the three remaining 2×2 matrices:

$$\begin{aligned} \det A &= 1 \cdot \begin{vmatrix} 4 & 8 \\ 6 & 9 \end{vmatrix} - 7 \cdot \begin{vmatrix} -4 & 8 \\ 2 & 9 \end{vmatrix} + 5 \cdot \begin{vmatrix} -4 & 4 \\ 2 & 6 \end{vmatrix} \\ &= (4 \cdot 9 - 8 \cdot 6) - 7((-4) \cdot 9 - 8 \cdot 2) + 5((-4) \cdot 6 - 4 \cdot 2) \\ &= -12 + 364 - 160 = 192 \end{aligned}$$

Next we shall evaluate the determinant by using the various properties to reduce the determinant to a triangular form.
By property (iii) we can extract the common factor of 4 from the second row:

$$\det A = 4 \begin{vmatrix} 1 & 7 & 5 \\ -1 & 1 & 2 \\ 2 & 6 & 9 \end{vmatrix}$$

and using corollary 3 we add the first row to the second one, and using corollary 3 again we subtract twice the first row from the last one, hence

$$\det A = 4 \begin{vmatrix} 1 & 7 & 5 \\ 0 & 8 & 7 \\ 0 & -8 & -1 \end{vmatrix}$$

then, using corollary 3 once more, we add the second row to the third one, hence

$$\det A = 4 \begin{vmatrix} 1 & 7 & 5 \\ 0 & 8 & 7 \\ 0 & 0 & 6 \end{vmatrix}$$

whence by property 3 we get $\det A = 4 \cdot 1 \cdot 8 \cdot 6 = 192$.

5.5) Determinant of block-diagonal, block-skew-diagonal and block-triangular matrices.
Let $A = (a_{ij})_{mm}$ and $B = (b_{ij})_{nn}$ be two square matrices and let

$$C = \begin{pmatrix} A & \mathcal{O} \\ \mathcal{O}^T & B \end{pmatrix} \tag{11}$$

where $\mathcal{O}$ is a null matrix. Then the following statement holds: The determinant of the block-diagonal matrix C is given by

$$\det C = \det A \det B \tag{12}$$

Indeed, $\det C$ is equal to the sum of products which are, up to a sign, equal to

$$a_{1j_1} a_{2j_2} \dots a_{mj_m} b_{1k_1} b_{2k_2} \dots b_{nk_n}$$

and the sum is carried out over all permutations of $[j]$ and all permutations of $[k]$. Let us keep $[j]$ fixed and carry out the summation over $[k]$, giving each term a sign according to $\varepsilon_{k_1k_2\ldots k_n}$. The latter sum is obviously equal to $\det B$. Then we carry out the sum over $[j]$, giving each factor the sign according to $\varepsilon_{j_1j_2\ldots j_m}$. As a result the a_{ij}-products add up to give $\det A$, and that sum is multiplied by $\det B$.

Example: Let C be the block-diagonal matrix

$$C = \begin{pmatrix} A & \mathcal{O} \\ \mathcal{O}^T & B \end{pmatrix}$$

with $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ and $B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$. Expanding $\det C$ about the first column and continuing to expand about the first column of each remaining determinant, we get

$$\begin{aligned} \det C &= a_{11} \begin{vmatrix} a_{22} & 0 & 0 & 0 \\ 0 & b_{11} & b_{12} & b_{13} \\ 0 & b_{21} & b_{22} & b_{23} \\ 0 & b_{31} & b_{32} & b_{33} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & 0 & 0 & 0 \\ 0 & b_{11} & b_{12} & b_{13} \\ 0 & b_{21} & b_{22} & b_{23} \\ 0 & b_{31} & b_{32} & b_{33} \end{vmatrix} \\ &= a_{11}a_{22} \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} - a_{21}a_{12} \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} = \det A \det B \end{aligned}$$

By similar reasoning we can convince ourselves that the following statements hold true:

$$\det C \equiv \begin{vmatrix} \mathcal{O} & A \\ B & \mathcal{O}^T \end{vmatrix} = -\det A \det B \tag{13}$$

and

$$\det C \equiv \begin{vmatrix} A & D \\ \mathcal{O} & B \end{vmatrix} = \det A \det B \tag{14}$$

where D and the null matrix $\mathcal{O}$ have appropriate dimensions.

5.6) Product of determinants.

> *The product of the determinants of matrices A and B is equal to the determinant of the matrix product AB:*
>
> $$\det A \det B = \det AB$$

The proof will be given first for 2×2 matrices. Then we shall present the general proof using block-triangular determinants.

Proof 1: Let $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ and $B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$, and consider the determinant

$$\begin{vmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ -1 & 0 & b_{11} & b_{12} \\ 0 & -1 & b_{21} & b_{22} \end{vmatrix}$$

We evaluate this determinant in two different ways: first by expansion about the first row, then by a sequence of elementary operations. The expansion about the first row gives

$$a_{11}\begin{vmatrix} a_{22} & 0 & 0 \\ 0 & b_{11} & b_{12} \\ -1 & b_{21} & b_{22} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & 0 & 0 \\ -1 & b_{11} & b_{12} \\ 0 & b_{21} & b_{22} \end{vmatrix}$$

and continuing expanding the 3×3 determinants about their first rows we get

$$\begin{aligned} a_{11}a_{22}\begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} - a_{12}a_{21}\begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} &= (a_{11}a_{22} - a_{12}a_{21})\begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} \\ &= \det A \det B \end{aligned}$$

Next we evaluate the determinant by first carrying out a sequence of elementary operations, aimed at getting zeros where the determinant has the matrix elements of A. Thus, add to the first row a_{11} times the third row and a_{12} times the fourth row, hence

$$\begin{vmatrix} 0 & 0 & a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21} & a_{22} & 0 & 0 \\ -1 & 0 & b_{11} & b_{12} \\ 0 & -1 & b_{21} & b_{22} \end{vmatrix}$$

and then add to the second row a_{21} times the third row and a_{22} times the fourth row, hence

$$\begin{vmatrix} 0 & 0 & a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ 0 & 0 & a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \\ -1 & 0 & b_{11} & b_{12} \\ 0 & -1 & b_{21} & b_{22} \end{vmatrix}$$

Now expand about the first column, hence

$$-\begin{vmatrix} 0 & a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ 0 & a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \\ -1 & b_{21} & b_{22} \end{vmatrix}$$

and expanding once more about the first column we get

$$\begin{vmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{vmatrix}$$

and we recognize that this is the determinant of the product AB, and hence the statement.

Proof 2: in terms of partitioned matrices.
Let A and B be two $n \times n$ matrices and let C be the block-triangular matrix

$$C = \begin{pmatrix} A & \mathcal{O} \\ -E & B \end{pmatrix}$$

where $\mathcal{O}$ is the $n \times n$ null matrix and E is the $n \times n$ unit matrix. From section 5.3.7 we know that $\det C = \det A \det B$. Now let us multiply the bottom row of $\det C$ by A and add it to the top row. This is an elementary operation which leaves the value of $\det C$ unchanged and gives it a skew-triangular form:

$$\det C = \begin{vmatrix} \mathcal{O} & AB \\ -E & B \end{vmatrix}$$

hence $\det C = \det(AB)$, i.e.

$$\det A \det B = \det(AB)$$

5.4) Minors and cofactors of a matrix; expansion theorem.
Definition: given the square matrix $A = (a_{ij})$, the minor M_{ij} of matrix element a_{ij} is the determinant, obtained from $\det A$ by crossing out the ith row and the jth column.

Example 6: Given the matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$$

the minor M_{13} is the following determinant :

$$M_{13} = \begin{vmatrix} a_{21} & a_{22} & a_{24} \\ a_{31} & a_{32} & a_{34} \\ a_{41} & a_{42} & a_{44} \end{vmatrix}$$

Definition: The cofactor A_{ij} of matrix element a_{ij} of A is defined by

$$A_{ij} = (-1)^{i+j} M_{ij}$$

We can now formulate the expansion theorem in full generality.
Theorem A: the determinant of matrix A is equal to the sum of the products of the matrix elements of the ith row with the corresponding cofactors, i.e.

$$\det A = \sum_{j=1}^{n} a_{ij} A_{ij} \tag{15}$$

Interesting is also the following theorem.
Theorem B: the sum of the products of the matrix elements of the ith row with the cofactors of the kth row, assuming $k \neq i$, is equal to nought:

$$\sum_{j=1}^{n} a_{ij} A_{kj} = 0, \qquad k \neq i$$

The two formulas, expressing theorems A and B, can be combined into one equation which is valid for any i and k:

$$\sum_{j=1}^{n} a_{ij} A_{kj} = \delta_{ik} \det A \tag{16}$$

where δ_{ij} is the Kronecker-δ symbol, $\delta_{ij} = \begin{cases} 1 & \text{if } k = i \\ 0 & \text{otherwise} \end{cases}$

Let us apply theorems A and B to Example 5:

$$A = \begin{pmatrix} 1 & 7 & 5 \\ -4 & 4 & 8 \\ 2 & 6 & 9 \end{pmatrix}$$

In the case of a 3×3 matrix there are nine cofactors:

$$\begin{array}{lcl} A_{11} & = & \begin{vmatrix} 4 & 8 \\ 6 & 9 \end{vmatrix} = -12, \quad A_{12} = -\begin{vmatrix} -4 & 8 \\ 2 & 9 \end{vmatrix} = 52, \quad A_{13} = \begin{vmatrix} -4 & 4 \\ 2 & 6 \end{vmatrix} - 32, \\ A_{21} & = & -\begin{vmatrix} 7 & 5 \\ 6 & 9 \end{vmatrix} = -33, \quad A_{22} = \begin{vmatrix} 1 & 5 \\ 2 & 9 \end{vmatrix} = -1, \quad A_{23} = -\begin{vmatrix} 1 & 7 \\ 2 & 6 \end{vmatrix} = 8, \\ A_{31} & = & \begin{vmatrix} 7 & 5 \\ 4 & 8 \end{vmatrix} = 36, \quad A_{32} = -\begin{vmatrix} 1 & 5 \\ -4 & 8 \end{vmatrix} = -28, \quad A_{33} = \begin{vmatrix} 1 & 7 \\ -4 & 4 \end{vmatrix} = 32, \end{array}$$

and we can check theorems A and B by explicit calculation, for instance:
(i)

$$\sum_{j=1}^{3} a_{2j} A_{2j} = -4 \cdot (-33) + 4 \cdot (-1) + 8 \cdot 8 = 192$$

(ii)

$$\sum_{j=1}^{3} a_{1j} A_{3j} = 1 \cdot 36 + 7 \cdot (-28) + 5 \cdot 32 = 0$$

and similarly one can check all remaining cases.

5.8) Inverse matrix.

In matrix algebra one defines the inverse A^{-1} of a square matrix A by the relation

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

where E is the unit matrix, and one shows that $A^{-1}A = AA^{-1}$ and that the inverse is unique. However, for the evaluation of A^{-1} one needs the concept of the determinant. This we have now introduced and studied in detail. We are therefore ready to proceed to the evaluation of the inverse matrix.

Thus assume that an $n \times n$ matrix A is given, and assume that a matrix X exists such that

$$AX = E \tag{17}$$

then, according to the definition of the inverse matrix, $X = A^{-1}$.

A first observation is that $\det(AX) = \det E = 1$. But we know from section 5.6 that $\det(AX) = \det A \det X$, and this implies that for A to have an inverse we must have $\det A \neq 0$. Such a matrix is called nonsingular; conversely a matrix whose determinant is equal to nought is called singular.

Let us re-write Eq.(17) in components:

$$\sum_j a_{ij} x_{jk} = \delta_{ik} \quad (i,\, k = 1,\, 2,\, \ldots,\, n)$$

If we keep k fixed and write the latter equation in detail for $i = 1, 2, \ldots, n$, then we get, setting for instance $k = 1$,

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n & = & 1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n & = & 0 \\ a_{31}x_1 + a_{32}x_2 + \ldots + a_{3n}x_n & = & 0 \\ \ldots \quad \ldots \quad \ldots \quad \ldots & & \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n & = & 0 \end{array}$$

where for simplicity we have omitted the second index of x. Now the latter equation is a simultaneous linear equation for the unknown x_j's. If we give k a different value, then the left-hand side of the equation remains unchanged, and on the right-hand side the 1 will appear in the equation that corresponds to k, all the other right-hand sides remaining equal to nought. This discussion shows that the evaluation of the inverse matrix is equivalent to solving a simultaneous linear equation. Let us do this in detail for the simple case of a 2×2 matrix.

Thus let $A = (a_{ij})_{22}$ with $\det A \neq 0$; we want to find the matrix $X = (x_{ij})_{22}$ that satisfies

$$AX = E \quad \text{or, in components,} \quad \sum_{j=1}^{2} a_{ij}x_{jk} = \delta_{ik}$$

For $k = 1$ we have

$$\begin{aligned} a_{11}x_{11} + a_{12}x_{21} &= 1 \\ a_{21}x_{11} + a_{22}x_{21} &= 0 \end{aligned}$$

and for $k = 2$

$$\begin{aligned} a_{11}x_{12} + a_{12}x_{22} &= 0 \\ a_{21}x_{12} + a_{22}x_{22} &= 1 \end{aligned}$$

Let us solve the equation for $k = 1$ by elimination. To eliminate x_{21} we multiply the upper equation by a_{22}, the lower equation by a_{12} and subtract, hence

$$(a_{11}a_{22} - a_{21}a_{12})x_{11} = a_{22}$$

and we recognise in the coefficient of x_{11} the determinant of A, hence

$$x_{11} = (\det A)^{-1}a_{22}$$

and similarly we get all elements of the inverse matrix. Thus the final result is

$$X \equiv A^{-1} = (\det A)^{-1} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$$

To verify our result we multiply A by X:

$$AX = (\det A)^{-1} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix} = E$$

We can re-write our final result in the following interesting form:

$$x_{ij} = \Delta^{-1}A_{ij}, \qquad i,\, j = (1,\, 2)$$

where $\Delta \equiv \det A$ and A_{ij} is the cofactor of a_{ij} (cf. section 5.7). This result remains true for matrices of any finite dimension. For a proof the reader is encouraged to consult a more substantial monograph on matrix algebra.

5.9) A note on the practical evaluation of determinants.

The numerical evaluation of 2×2 and 3×3 determinants is straight forward, but already 4×4 determinants are tedious, and higher order determinants pose formidable computational problems. One of the difficulties lies in the alternating signs of the sum of products that define the determinant. Assuming that all matrix elements are of equal order of magnitude and, for the sake of the argument, that they are all positive, one can immediately see that all terms of the determinant are of similar order of magnitude, half of them positive, the other half negative, and as a result the most significant digits cancel. If the matrix elements are decimal fractions, it can easily happen that what is left after the cancellation is dominated by rounding errors and bears no resemblance to the true value of the determinant. Therefore the evaluation of determinants is a problem for specialists in numerical analysis, who have developed sophisticated algorithms. They take account of peculiarities of determinants. Thus one considers separately sparse determinants, i.e. determinants whose elements are mostly zero, or band-diagonal determinants, i.e. determinants whose principal diagonal and a few adjacent diagonals have nonzero elements and zeros otherwise. Such algorithms are implemented in computer programs available in packages, such as "*Numerical Recipes:* the Art of Scientific Computing" by W.H. Press et al.

5.10. Exercise.
Show that

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ 2 & 2^2 & 2^3 & 2^4 \\ 3 & 3^2 & 3^3 & 3^4 \\ 4 & 4^2 & 4^3 & 4^4 \end{vmatrix} = 2! \cdot 3! \cdot 4! \tag{18}$$

Hint: subtract the third column from the fourth, then the second column from the third, the first from the second, then expand about the first row to get a 3×3 determinant; extract all common factors, hence show by further elementary operations that the remaining determinant is equal to 2. Note: It is generally true that

$$\begin{vmatrix} 1 & 1 & 1 & \dots & 1 \\ 2 & 2^2 & 2^3 & \dots & 2^n \\ 3 & 3^2 & 3^3 & \dots & 3^n \\ \dots & \dots & \dots & \dots & \dots \\ n & n^2 & n^3 & \dots & n^n \end{vmatrix} = 2! \cdot 3! \cdot 4! \dots n! \tag{19}$$

6. Trace of a matrix.

Definition: The *trace* of a matrix is defined as the sum of its diagonal elements.
This definition implies that the concept of the trace applies only to square matrices. There is no universal notation for the trace, but in the physics literature one uses mostly $\mathrm{Tr}\, A$ to denote the trace of the matrix A.

Example: The trace of $A = (a_{ij})_{nn}$ is

$$\mathrm{Tr}\, A = \sum_{i=1}^{n} a_{ii}$$

In particular the trace of the unit $n \times n$ matrix is equal to n.

We note the following properties of the trace:

(i) If A and B are two $n \times n$ matrices, then

$$\mathrm{Tr}\,(A + B) = \mathrm{Tr}\, A + \mathrm{Tr}\, B$$

(ii) If A is an $n \times n$ matrix and c is a number, then

$$\mathrm{Tr}\,(cA) = c\mathrm{Tr}\, A$$

(iii) If A and B are two matrices of such dimensions that both products AB and BA exist and are square matrices, then

$$\mathrm{Tr}\,(AB) = \mathrm{Tr}\,(BA)$$

The proofs of properties (i) and (ii) can be done easily from the definition of the trace. The proof of property (iii) is slightly more involved but also not particularly difficult or lengthy, but rather than giving the general proof let us consider a particular example: let

$$A = \begin{pmatrix} 1 & -1 & 2 \\ -1 & 3 & 1 \end{pmatrix} \qquad \text{and} \qquad B = \begin{pmatrix} 4 & 2 \\ -5 & 7 \\ -1 & 3 \end{pmatrix}$$

hence

$$AB = \begin{pmatrix} 7 & 1 \\ -20 & 22 \end{pmatrix}, \qquad BA = \begin{pmatrix} 2 & 2 & 10 \\ -12 & 26 & -3 \\ -4 & 10 & 1 \end{pmatrix}$$

and hence $\mathrm{Tr}\,(AB) = \mathrm{Tr}\,(BA) = 29$.

It must be strongly emphasised that the identity $\mathrm{Tr}\,(AB) = \mathrm{Tr}\,(BA)$ does *not* depend on the two matrix products, AB and BA, being of equal dimensions; all that is required is that the two products exist and are square matrices. This condition is of course always satisfied if A and B are themselves square matrices of equal dimension.

7. Linear independence of matrices.

Consider the set matrices of equal dimensions A_1, A_2, ..., A_n. The expression

$$c_1 A_1 + c_2 A_2 + \ldots + c_n A_n,$$

where c_1, c_2, ..., c_n are arbitrary numbers, is called a *linear superposition* of the matrices A_i.

The set of matrices A_i is *linearly dependent* if we can find nonzero values of the superposition coefficients c_i such that

$$c_1 A_1 + c_2 A_2 + \ldots + c_n A_n = 0 \tag{20}$$

If no such set of superpositions exists, i.e. if Eq. (20) holds *only* for $c_i = 0$, $i = 1, 2, \ldots, n$, then the matrices A_i are said to be *linearly independent*.

Example: consider the set of 2×2 matrices

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

(Pauli matrices).[2] Let us form their linear superposition and set it equal to nought:

$$\Sigma \equiv c_0\sigma_0 + c_1\sigma_1 + c_2\sigma_2 + c_3\sigma_3 = 0$$

To decide whether the Pauli matrices are linearly independent we can procede like this: we take the trace of Σ and use the properties of the trace from the preceeding section:

$$\mathrm{Tr}\,\Sigma = \mathrm{Tr}\,(c_0\sigma_0 + c_1\sigma_1 + c_2\sigma_2 + c_3\sigma_3) = c_0\mathrm{Tr}\,\sigma_0 + c_1\mathrm{Tr}\,\sigma_1 + c_2\mathrm{Tr}\,\sigma_2 + c_3\mathrm{Tr}\,\sigma_3 = 2c_0 = 0$$

where we have used $\mathrm{Tr}\,\sigma_0 = 2$ and $\mathrm{Tr}\,\sigma_i = 0$ for $i = 1, 2, 3$, which can be seen by inspection of the Pauli matrices. Thus $c_0 = 0$.

[2] usually only the set σ_1, σ_2, σ_3 are called Pauli matrices; I am differing from that convention to simplify the following discussion.

Next we multiply Σ by σ_1 and use the property of the Pauli matrices that $\sigma_i\sigma_j = \delta_{ij}\sigma_0 + i\sum_{k=1}^{3}\varepsilon_{ijk}\sigma_k$, and then take the trace, hence

$$\mathrm{Tr}\,\sigma_1\Sigma = c_0\mathrm{Tr}\,\sigma_1 + c_1\mathrm{Tr}\,(\sigma_1)^2 + c_2\mathrm{Tr}\,\sigma_1\sigma_2 + c_3\mathrm{Tr}\,\sigma_1\sigma_3 = 2c_1 = 0$$

i.e. $c_1 = 0$. Similarly we show that $c_2 = 0$ and $c_3 = 0$. Therefore the four Pauli matrices are linearly independent.

8. Special matrices.

8.1) Symmetric and antisymmetric matrices.
A matrix $A = (a_{ij})$ is called *symmetric* if $A^T = A$. Expressed in terms of the matrix elements the symmetry condition is $a_{ji} = a_{ij}$.

Example: The following matrices are symmetric:

$$\begin{pmatrix} 2 & 4 & -1 \\ 4 & 3 & 7 \\ -1 & 7 & 9 \end{pmatrix}, \qquad \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}$$

A matrix $A = (a_{ij})$ is called *antisymmetric* if $A^T = -A$. Expressed in terms of the matrix elements the symmetry condition is $a_{ji} = -a_{ij}$.

Example: The following matrices are antisymmetric:

$$\begin{pmatrix} 0 & 4 & -1 \\ -4 & 0 & 7 \\ 1 & -7 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & a_{12} & a_{13} \\ -a_{12} & 0 & a_{23} \\ -a_{13} & -a_{23} & 0 \end{pmatrix}$$

It should be noted that the diagonal elements of antisymmetric matrices are equal to nought. This follows directly from the condition $a_{ji} = -a_{ij}$ if applied to the diagonal elements: $a_{ii} = -a_{ii}$, hence $a_{ii} = 0$.

Of interest is also the following statement: an arbitrary $n \times n$ matrix M can be uniquely represented as the sum of a symmetric matrix S and an antisymmetric matrix A.
Proof: Let $M = S + A$, where $S^T = S$ and $A^T = -A$. Then $M^T = S^T + A^T = S - A$, and hence $S = \frac{1}{2}(M + M^T)$ and $A = \frac{1}{2}(M - M^T)$.

8.2) Orthogonal matrix.
A matrix $A = (a_{ij})$ is called *orthogonal* if

$$A^T A = AA^T = E \tag{21}$$

Expressed in terms of the matrix elements the orthogonality condition is

$$\sum_{i=1}^{n} a_{ij}a_{ik} = \delta_{jk} \tag{22}$$

The definition implies that A is a square matrix. Therefore $\det A$ and $\det A^T$ exist and are equal. We have therefore

$$\det AA^T = \det A \det A^T = (\det A)^2 = \det E = 1$$

hence $\det A = \pm 1$.

The product of two orthogonal matrices A and B is an orthogonal matrix. Indeed, we have

$$\begin{aligned} AB(AB)^T = ABB^TA^T &= AA^T \quad \text{by the orthogonality of } B \\ &= E \qquad \text{by the orthogonality of } A \end{aligned} \tag{23}$$

If we consider the orthogonal matrix in a form partitioned by columns and call each column a vector, then we can use the following notation:

$$A = \left(\vec{a}^{(1)}\, \vec{a}^{(2)}\, \ldots\, \vec{a}^{(n)}\right)$$

and the orthogonality relation (22) can be written in the following form:

$$\vec{a}^{(i)}\vec{a}^{(j)} = \delta_{ij}$$

which means that the column vectors $\vec{a}^{(i)}$ are normalized and mutually orthogonal. Similarly one can show that the rows of orthogonal matrices are normalized and mutually orthogonal.

Now recall the expansion theorem of determinants, Eq. (15) or (16):

$$\det A = \sum_{j=1}^{n} a_{ij}A_{ij}, \qquad \sum_{j=1}^{n} a_{ij}A_{kj} = \delta_{ik} \det A \tag{24}$$

where A_{ij} is the cofactor of a_{ij}. If we compare this with the orthonality relation (22), then we can conclude that the cofactors A_{ij} of an orthogonal matrix are equal to the corresponding matrix elements a_{ij} up to a sign:

$$A_{ij} = \text{sign}(\det A) a_{ij} \tag{25}$$

We shall need this important property of orthogonal matrices for the proof that the cross product of two vectors is itself a vector.

Orthogonal matrices are particularly important in connection with coordinate transformations. Indeed, consider the vector $\vec{r} = (x, y)$. We have previously established that a rotation of the coordinates through an angle α, leaving the vector itself unchanged, gives new components of $\vec{r}$, denoted x' and y', which are related to x and y by

$$x' = x\cos\alpha + y\sin\alpha, \qquad y' = -x\sin\alpha + y\cos\alpha$$

In matrix notation this transformation can be written as

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

and we can easily check that the matrix $\Omega = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}$ is an orthogonal matrix.

We can extend the argument to three dimensions and indeed to n dimensional vector spaces. In the three-dimensional case we define the following orthogonal matrices:

$$R_x(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c\alpha & s\alpha \\ 0 & -s\alpha & c\alpha \end{pmatrix}, \quad R_y(\beta) = \begin{pmatrix} c\beta & 0 & s\beta \\ 0 & 1 & 0 \\ -s\beta & 0 & c\beta \end{pmatrix}, \quad R_z(\gamma) = \begin{pmatrix} c\gamma & s\gamma & 0 \\ -s\gamma & c\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

where $c\alpha = \cos\alpha$, $s\alpha = \sin\alpha$, $c\beta = \cos\beta, \ldots, s\gamma = \sin\gamma$. The angles α, β and γ are called Euler angles. The Euler rotations are of interest because they allow us to write down a rotation in three-dimensional space about an arbitrary axis A through an arbitrary angle θ. To do this one must carry out a sequence of three rotations:

(i) a rotation about the z axis to bring axis A into the (new) yz plane,

(ii) a rotation about the new x axis through an angle β to make the z axis coincide with A,
and finally

(iii) a rotation about the z axis through the angle γ.

In symbols we can write this in the following form:

$$(xyz) \stackrel{R_z(\alpha)}{\longrightarrow} (x'y'z') \stackrel{R_x(\beta)}{\longrightarrow} (x''y''z'') \stackrel{R_z(\gamma)}{\longrightarrow} (x'''y'''z''') \tag{26}$$

This sequence of rotations is expressed by the matrix product $R(\alpha,\beta,\gamma) = R_z(\gamma)R_x(\beta)R_z(\alpha)$. The orthogonality of R_x and R_z ensures the orthogonality of R:

$$\begin{aligned} R^T R &= [R_z(\gamma)R_x(\beta)R_z(\alpha)]^T \; R_z(\gamma)R_x(\beta)R_z(\alpha) \\ &= R_z^T(\alpha)R_x^T(\beta)R_z^T(\gamma)R_z(\gamma)R_x(\beta)R_z(\alpha) = 1 \end{aligned}$$

We can now prove several statements concerning transformations of vectors, which were left without proof in the chapter on vector algebra:

(i) the length of a vector $\vec{a}$ is invariant under rotations,

(ii) the scalar product of vectors $\vec{a}$ and $\vec{b}$ is invariant under rotations,

(iii) the cross product of vectors $\vec{a}$ and $\vec{b}$ transforms like a vector.

Note that (i) and (ii) are logically the same statement. Indeed, the square of the length of $\vec{a}$ is given by $\vec{a}\cdot\vec{a}$, i.e. by the scalar product of $\vec{a}$ with itself. It is therefore sufficient to prove (ii): Let the vectors after the rotation be $\vec{a}\,'$ and $\vec{b}\,'$, respectively, i.e.

$$\vec{a}\,' = R\vec{a} \qquad \text{and} \quad \vec{b}\,' = R\vec{b}$$

then

$$\vec{a}\,'\cdot\vec{b}\,' \equiv \sum_i a_i' b_i' = \sum_i (R\vec{a})_i (R\vec{b})_i == \sum_i \left(\sum_j R_{ij}a_j\right)\left(\sum_k R_{ik}b_k\right) = \sum_{jk}\left(\sum_i R_{ij}R_{ik}\right) a_j b_k$$

and by Eq. (21) or (22) we have $\left(\sum_i R_{ij}R_{ik}\right) = \delta_{jk}$, hence

$$\vec{a}\,'\cdot\vec{b}\,' = \sum_{jk}\delta_{jk}a_j a_k = \sum_j a_j b_j = \vec{a}\cdot\vec{b}$$

which is the statement of rotational invariance of $\vec{a} \cdot \vec{b}$.

Finally let us prove the statement that the cross product $\vec{c}$ of vectors $\vec{a} = (a_1, a_2, a_3)$ and $\vec{b} = (b_1, b_2, b_3)$ is itself a vector. Thus let

$$\vec{c} = \vec{a} \times \vec{b}$$

and let $\vec{a}$ and $\vec{b}$ transform under a rotation R according to

$$a_i \xrightarrow{R} a'_i = R_{ij} a_j, \qquad b_i \xrightarrow{R} b'_i = R_{ij} b_j$$

where summation over j is implied. To prove that $\vec{c}$ is a vector we must show that $c_i \xrightarrow{R} c'_i = R_{ij} c_j$. Now we have

$$c_i \xrightarrow{R} c'_i = \left(\vec{a}' \times \vec{b}'\right)_i = \varepsilon_{ijk}\, a'_j\, b'_k = \varepsilon_{ijk}\, (R_{jl}\, a_l)\, (R_{km}\, b_m) = (\varepsilon_{ijk}\, R_{jl} R_{km})\, a_l\, b_m$$

Here in the expression $\varepsilon_{ijk}\, R_{jl} R_{km}$ summations over j and k are implied; therefore this expression is a third-rank tensor, which we can denote by A_{ilm}. Let us establish that $A_{imm} = 0$. To do this consider the particular case A_{1mm}:

$$A_{1mm} = \varepsilon_{1jk} R_{jm} R_{km} = \varepsilon_{123} R_{2m} R_{3m} + \varepsilon_{132} R_{3m} R_{2m} = 0$$

where in the last step we have used $\varepsilon_{ikj} = -\varepsilon_{ijk}$. The cases A_{2mm} and A_{3mm} can be done similarly.

Next let us show that A_{ijk} is antisymmetric in the indices (jk). Indeed, using the antisymmetry of ε_{ijk} we have

$$A_{ilm} = \varepsilon_{ijk} R_{jl} R_{km} = -\varepsilon_{ikj} R_{jl} R_{km} = -A_{iml}$$

Now consider the particular case of A_{123}:

$$A_{123} = \varepsilon_{1jk} R_{2j} R_{3k} = R_{22} R_{33} - R_{23} R_{32}$$

hence, with Eq. (25) and noting that $\det R = +1$, we get $A_{123} = R_{11}$. Similarly we can find all other nonzero elements of A_{ijk}, and hence, collecting our results, we find

$$A_{ilm} = R_{ij} \varepsilon_{jlm}$$

hence

$$c'_i = A_{ilm} a_l b_m = R_{ij} \varepsilon_{jlm} a_l b_m = R_{ij} c_j$$

which is the correct transformation for a vector.

8.3) Hermitian and antihermitian matrices.
The matrix $H = (h_{ij})$ is called hermitian if it has the property

$$H^\dagger = H$$

where the † symbol denotes hermitian conjugation (cf. section 2). This definition implies that a hermitian matrix is a square matrix. Written in components, the definition of the hermitian matrix can be restated thus:

$$h^*_{ji} = h_{ij}$$

where the asterisk denotes complex conjugation. In particular, we have for the diagonal elements $h_{ii}^* = h_{ii}$, i.e. the diagonal elements of a hermitian matrix are real.

The property of hermiticity is an extension of symmetry to complex matrices. We can also define the complex analogue of an antisymmetric matrix, namely an *antihermitian* matrix:

A matrix A is called antihermitian if it has the property

$$A^\dagger = -A$$

From a discussion similar to the one that taught us of the reality of the diagonal elements of a hermitian matrix, we can see that the antihermitian matrix has imaginary diagonal elements.

Interesting is the following statement:
Any complex matrix C can be represented as the sum of an hermitian matrix H and an antihermitian matrix A; H and A are uniquely defined.

Indeed, let $C = H + A$, then $C^\dagger = H^\dagger + A^\dagger = H - A$. If we add these two equations, we get $H = \frac{1}{2}(C + C^\dagger)$, and if we subtract them we get $A = \frac{1}{2}(C - C^\dagger)$.

We can also write an antihermitian matrix A as $A = iH$, where H is hermitian. Therefore the separation of a complex matrix C into hermitian and antihermitian matrices can be written in the form of

$$C = H_1 + iH_2$$

where H_1 and H_2 are hermitian matrices. In this form one recognises the analogue of the familiar representation of a complex number as the sum of its real part and i-times its imaginary part.

Another important property of hermitian matrices will be discussed in section 9, namely that the eigenvalues of hermitian matrices are real.

8.4) Unitary matrix.
A complex matrix U is called unitary if it has the property

$$UU^\dagger = E \tag{27}$$

where E is a unit matrix.

Note that, if U is a real matrix, then hermitian conjugation is the same as transposition, and the above definition of the unitary matrix coincides with that of an orthogonal matrix. Thus unitarity is the generalization of orthogonality to complex matrices.

Using the property of determinants that $\det(AB) = \det(A)\det(B)$ (cf. section 5.6), we can see that

$$\det(U)\det(U^\dagger) = 1$$

This implies that a unitary matrix is regular and therefore its inverse U^{-1} exists. If we then premultiply Eq. (27) by U^{-1}, we get

$$U^\dagger = U^{-1}$$

and if we postmultiply this by U we get

$$U^\dagger U = E$$

Comparing this with Eq. (27) we conclude that a unitary matrix commutes with its hermitian conjugate.

9. Eigenvalues and eigenvectors; diagonalization of matrices.

9.1) In this section we refer to a column matrix as *vector* irrespective of any transformation properties that were considered as the essential property of any vector in vector algebra. Thus we shall denote a vector by $\vec{x}$ or $\vec{y}$ and the corresponding components by x_i and y_i, $i = 1, 2, \ldots, n$, such that

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \qquad \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

9.2) The product of a vector $\vec{x}$, say, by the $n \times n$ matrix $A = (a_{ij})$ is a vector:

$$A\vec{x} = \vec{y} \tag{28}$$

Generally speaking the vectors $\vec{x}$ and $\vec{y}$ have different directions in n dimensional vector space. Only exceptionally $\vec{y}$ will be collinear with $\vec{x}$, i.e. $\vec{y} = \lambda\vec{x}$, and we can write instead of Eq. (28):

$$A\vec{x} = \lambda\vec{x} \tag{29}$$

A vector that satisfies Eq. (29) is called an *eigenvector* of A and λ is called the corresponding *eigenvalue*. We shall see that in general a matrix has several eigenvalues and eigenvectors. The problem of finding all eigenvalues and eigenvectors of a matrix is called the *eigenvalue problem.*[3]

Of particular interest in physical applications is the eigenvalue problem of symmetric and, more generally, of Hermitian matrices. We shall therefore have these matrices in mind in the following discussion.

9.3) Thus consider Eq. (29), where A is a known $n \times n$ matrix and $\vec{x}$ and λ are the unknown eigenvector and eigenvalue which are to be determined. An equivalent form of the eigenvalue equation is

$$(A\vec{x} - \lambda E)\vec{x} = 0 \tag{30}$$

where E is the $n \times n$ unit matrix. Writing the latter equation explicitly we have

$$\begin{pmatrix} a_{11}-\lambda & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & a_{22}-\lambda & a_{23} & \ldots & a_{1n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & a_{n3} & \ldots & a_{nn}-\lambda \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \ldots \\ 0 \end{pmatrix} \tag{31}$$

or, carrying out the multiplication on the l.h.s. and equating every component of the resulting vector to the corresponding component of the null vector on the r.h.s., we get

$$\begin{aligned} (a_{11}-\lambda)x_1 + a_{12}x_2 + a_{13}x_3 + \ldots + a_{1n}x_n &= 0 \\ a_{21}x_1 + (a_{22}-\lambda)x_2 + a_{23}x_3 + \ldots + a_{1n}x_n &= 0 \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots & \\ a_{n1}x_1 + a_{n2}x_2 + a_{n3}x_3 + \ldots + (a_{nn}-\lambda)x_n &= 0 \end{aligned} \tag{32}$$

In the last form, Eq. (32), we recognize a linear simultaneous equation with n unknowns, the x_i, $i = 1, 2, \ldots, n$. This equation has the obvious solution $x_1 = x_2 = \ldots = x_n = 0$, but this

[3]strictly we should say "eigenvalue problem of matrix algebra" since there are also eigenvalue problems in other branches of mathematics, for instance in the theories of differential and of integral equations.

solution is of no interest. It is called the *trivial* solution. There exist also *nontrivial* solutions, and it is these that we now proceed to find.

9.4) Consider for simplicity the two-dimensional case, i.e. $n = 2$. We can then write Eq. (32) in the form of

$$(a_{11} - \lambda)x_1 + a_{12}x_2 = 0 \tag{33}$$
$$a_{21}x_1 + (a_{22} - \lambda)x_2 = 0 \tag{34}$$

These two equations are the equations of straight lines through the origin of the (x_1, x_2) plane (see Fig. 1a). The slopes of these lines are given by

$$x_2 = -\frac{a_{11} - \lambda}{a_{12}}x_1 \qquad \text{and} \qquad x_2 = -\frac{a_{21}}{a_{22} - \lambda}x_1 \tag{35}$$

respectively, and generally speaking they will be different. In this case the two lines have only one point in common, namely the origin, and only the trivial solution exists. However, if we can find a value of λ such that the slopes are equal, then the lines coincide (see Fig. 1b) and we get also nontrivial solutions. Such a value of λ is called an *eigenvalue* of A. If a nontrivial solution is found then we can also find other nontrivial solutions by multiplying x_1 and x_2 by the same constant, since only the ratio of x_1 to x_2 is fixed by the equations. Any pair (x_1, x_2) that satisfies simultaneously Eqs. (33) and (34) is called an *eigenvector* of A.

9.5) Let us now proceed to find the eigenvalues of A. Considering Eq. (35) we see that the condition of equality of the slopes of the two lines is given by

$$\frac{a_{11} - \lambda}{a_{12}} = \frac{a_{21}}{a_{22} - \lambda} \qquad \text{or} \qquad (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} = 0 \tag{36}$$

Especially in the latter form of Eq. (36) do we recognize that the l.h.s. is the determinant of $A - \lambda E$. We can therefore rewrite Eq. (36) in the equivalent form of

$$\det(A - \lambda E) = 0 \tag{37}$$

Considered as an equation in the unknown λ, Eq. (36) is a quadratic equation which we can also write in the expanded form of

$$\lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21} = 0 \tag{38}$$

and we see that the coefficient of λ is, up to a sign, the trace of A and the constant term is the determinant of A. Thus, if we set $T \equiv \operatorname{Tr} A = a_{11} + a_{22}$ and $\Delta \equiv \det A = a_{11}a_{22} - a_{12}a_{21}$, then we can cast Eq. (38) in the form of

$$\lambda^2 - T\lambda + \Delta = 0 \tag{39}$$

Equations (36) to (39) are all equivalent; they are called the *characteristic equation* of A.

The roots of the characteristic equation are

$$\lambda_{1,2} = \frac{1}{2}T \pm \sqrt{\frac{1}{4}T^2 - \Delta} \tag{40}$$

and in general we can have one of three cases:
(i) the two roots are real and distinct if $\frac{1}{4}T^2 - \Delta > 0$;

(ii) there is one double root if $\frac{1}{4}T^2 - \Delta = 0$, and
(iii) there is a pair of complex conjugate roots if $\frac{1}{4}T^2 - \Delta < 0$.

9.6) Up to this point in the derivation it was immaterial whether A was a Hermitian matrix or not. But this will be important now, because we shall prove quite generally that the eigenvalues of a Hermitian matrix are real.

Indeed, a Hermitian matrix has the property that its diagonal elements are real and the off-diagonal elements a_{ij} are related by $a_{ij} = a^*_{ji}$. Thus in the particular case of a Hermitian 2×2 matrix we have $a_{12} = a^*_{21}$, and hence $a_{12}a_{21} = |a_{12}|^2$, and therefore we get

$$\frac{1}{4}T^2 - \Delta = \frac{1}{4}(a_{11} + a_{22})^2 - (a_{11}a_{22} - |a_{12}|^2) = \frac{1}{4}(a_{11} - a_{22})^2 + |a_{12}|^2 > 0 \qquad (41)$$

Thus the characteristic equation of a Hermitian 2×2 matrix has only real roots.

9.7) We now proceed to prove that all eigenvalues of a Hermitian matrix of any order are real. To do this we go back to Eq. (29) and consider it together with its Hermitian conjugate:

$$A\vec{x} = \lambda\vec{x}, \qquad \vec{x}^\dagger A = \lambda^* \vec{x}^\dagger \qquad (42)$$

where in the latter equation we have used the definition of the Hermitian matrix, $A^\dagger = A$. If we now premultiply the first of Eqs. (42) by $\vec{x}^\dagger$ and postmultiply the second equation by $\vec{x}$ and then subtract, then we get

$$(\lambda - \lambda^*)\vec{x}^\dagger\vec{x} = 0$$

and since $\vec{x}^\dagger\vec{x} \neq 0$ it follows that

$$\lambda - \lambda^* = 0$$

i.e. the eigenvalue λ is real.

One can also prove the converse theorem: if a matrix has only real eigenvalues, then it is Hermitian.

Note that our proof makes no reference to the dimension of A. It is therefore valid for any $n \times n$ Hermitian matrix, indeed also for infinite dimensional Hermitian matrices, which are used in quantum mechanics.

9.8) To study the properties of the characteristic equation further, let us extend the discussion from the two-dimensional to the three-dimensional case. The characteristic equation of a 3×3 matrix is

$$\begin{aligned}\det(A - \lambda E) &= \begin{vmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ a_{21} & a_{22} - \lambda & a_{23} \\ a_{31} & a_{32} & a_{33} - \lambda \end{vmatrix} \\ &= -\lambda^3 + T\lambda^2 - M\lambda + \Delta = 0 \end{aligned} \qquad (43)$$

where $T = \mathrm{Tr}\, A = a_{11} + a_{22} + a_{33}$, $\Delta = \det A$, $M = C_{11} + C_{22} + C_{33}$, and C_{ii} are the cofactors of the diagonal elements of A:

$$C_{11} = a_{22}a_{33} - a_{23}a_{32}, \qquad C_{22} = a_{11}a_{33} - a_{13}a_{31}, \qquad C_{33} = a_{11}a_{22} - a_{12}a_{21}$$

Thus for the three-dimensional case the characteristic equation is a cubic equation in λ. Similarly we can find that in the general n dimensional case the characteristic equation is of nth order:

$$\det(A - \lambda E) = (-\lambda)^n + \mathrm{Tr}\, A\lambda^{n-1} + \ldots + \Delta = 0 \qquad (44)$$

Assuming that the characteristic equation has n distinct roots λ_i, we can also write

$$\begin{aligned}\det(A-\lambda E) &= (\lambda_1-\lambda)(\lambda_2-\lambda)\ldots(\lambda_n-\lambda) \\ &= (-\lambda)^n+(\lambda_1+\lambda_2+\ldots+\lambda_n)\lambda^{n-1}+\ \ldots\ +\lambda_1\lambda_2\ldots\lambda_n=0 \qquad (45)\end{aligned}$$

and comparing with Eq. (44) we find the interesting result that the sum of the eigenvalues is equal to the trace of A and their product is equal to the determinant of A.

9.9) To solve the eigenvalue problem completely we must also find the eigenvectors of A. We shall again do the derivation in detail in two dimensions before dealing with the general n-dimensional case.

It will be very important now to distinguish the cases when all roots are distinct and when some or all of them are multiple roots. In the latter case one says that the eigenvalues are *degenerate*. We shall deal with degeneracy in section 9.13. Here we assume that all eigenvalues are distinct, i.e. that they are *nondegenerate*.

For the two-dimensional case we go back to the simultaneous equations (33) and (34) or to the equivalent equations (35), where we now assume that λ is a known eigenvalue. Thus, substituting the first eigenvalue, λ_1, we find for the components of the eigenvector

$$x_2^{(1)} = -\frac{a_{11}-\lambda_1}{a_{12}}x_1^{(1)}$$

where we have indicated by the superscripts to which eigenvalue this eigenvector belongs.

We can see that the latter equation determines only the *ratio* of the components. This was to be expected from the discussion of the geometrical meaning of the eigenvectors given in section 4 (cf. Fig. 1b). It is therefore usual to define the eigenvectors such that they are **normalized**, i.e. that their norm is equal to one. This is done by first assigning an arbitrary value to one of the components, for instance to set $x_1^{(1)}=1$ and to write the eigenvector in the form

$$\vec{x}^{(1)} = N\begin{pmatrix}1\\(\lambda_1-a_{11})/a_{12}\end{pmatrix}$$

where N is a normalization factor which is determined by the condition

$$|\vec{x}|^2 \equiv \vec{x}^\dagger\vec{x} = |N|^2(|x_1^{(1)}|^2+|x_2^{(1)}|^2)=1$$

Similarly we get the second eigenvector by substituting λ_2 in Eq. (35) and proceeding as before.

Example. Given the Hermitian matrix $\sigma=\begin{pmatrix}0&-i\\i&0\end{pmatrix}$ we want to find its eigenvalues and eigenvectors.
The characteristic equation is

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

hence $\lambda_{1,2}=\pm1$. The ratios of the components of the eigenvectors are given by

$$x_2 = i\lambda x_1$$

and we get for the eigenvectors

$$\vec{x}^{(1)}=N_1\begin{pmatrix}1\\i\end{pmatrix},\qquad \vec{x}^{(2)}=N_2\begin{pmatrix}1\\-i\end{pmatrix}$$

and applying the normalization condition we get

$$\vec{x}^{(1)\dagger}\vec{x}^{(1)} = |N_1|^2(1 \ -i)\begin{pmatrix} 1 \\ i \end{pmatrix} = 2|N_1|^2 = 1$$

and we see that the phase of N_1 remains undetermined. We can choose it arbitrarily, and for convenience we shall make N_1 real and positive, i.e. we set $N_1 = 1/\sqrt{2}$. Similarly we find that $N_2 = 1/\sqrt{2}$. Thus finally we have the normalized eigenvectors of σ:

$$\vec{x}^{(1)} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix}, \qquad \vec{x}^{(2)} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}$$

9.10) It is interesting to note that the two eigenvectors are *orthogonal*, which means that

$$\vec{x}^{(1)\dagger}\vec{x}^{(2)} = x_1^{(1)*}x_1^{(2)} + x_2^{(1)*}x_2^{(2)} = 0$$

This is a particular case of a general property of eigenvectors:

Eigenvectors belonging to different eigenvalues are orthogonal.

To prove this let us denote two eigenvalues of A by λ_1 and λ_2 and the corresponding eigenvectors by $\vec{x}^{(1)}$ and $\vec{x}^{(2)}$. Then we have the two identities

$$A\vec{x}^{(1)} = \lambda_1\vec{x}^{(1)}$$

and

$$A\vec{x}^{(2)} = \lambda_2\vec{x}^{(2)}$$

Take the Hermitian conjugate of the latter equation, use the Hermiticity of A and the reality of λ, hence

$$\vec{x}^{(2)\dagger}A = \lambda_2\vec{x}^{(2)\dagger}$$

postmultiply this equation by $\vec{x}^{(1)}$, premultiply the equation for λ_1 by $\vec{x}^{(2)\dagger}$ and subtract. As a result we get

$$0 = (\lambda_1 - \lambda_2)\vec{x}^{(2)\dagger}\vec{x}^{(1)}$$

and since by assumption $\lambda_1 - \lambda_2 \neq 0$ we have the result that $\vec{x}^{(2)\dagger}\vec{x}^{(1)} = 0$.

9.11) An interesting consequence of the orthonormality of the normalized eigenvectors is the following:

> *if we construct a matrix, whose columns are the normalized eigenvectors of the Hermitian matrix A, then this matrix is unitary.*

Let us denote the matrix, constructed from the eigenvectors of A, by X, i.e.

$$X = \begin{pmatrix} x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(n)} \\ x_2^{(1)} & x_2^{(2)} & \dots & x_2^{(n)} \\ \dots & \dots & \dots & \dots \\ x_n^{(1)} & x_n^{(2)} & \dots & x_n^{(n)} \end{pmatrix} \tag{46}$$

which can be written also in a more compact form as a partitioned matrix, partitioned in columns, each column represented by the corresponding vector:

$$X = \left(\vec{x}^{(1)} \quad \vec{x}^{(2)} \quad \dots \quad \vec{x}^{(n)} \right) \tag{47}$$

If we multiply X by its Hermitian conjugate and use the orthonormality of the eigenvectors, we get

$$X^\dagger X = \begin{pmatrix} \vec{x}^{(1)\dagger} \\ \vec{x}^{(2)\dagger} \\ \dots \\ \vec{x}^{(n)\dagger} \end{pmatrix} \left(\vec{x}^{(1)} \quad \vec{x}^{(2)} \quad \dots \quad \vec{x}^{(n)} \right) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix} \tag{48}$$

which means that X is a unitary matrix (or an orthogonal matrix if it is real).

If we premultiply X by A, then every column is multiplied by the corresponding eigenvalue:

$$AX = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(n)} \\ x_2^{(1)} & x_2^{(2)} & \dots & x_2^{(n)} \\ \dots & \dots & \dots & \dots \\ x_n^{(1)} & x_n^{(2)} & \dots & x_n^{(n)} \end{pmatrix} = \begin{pmatrix} \lambda_1 x_1^{(1)} & \lambda_2 x_1^{(2)} & \dots & \lambda_n x_1^{(n)} \\ \lambda_1 x_2^{(1)} & \lambda_2 x_2^{(2)} & \dots & \lambda_n x_2^{(n)} \\ \dots & \dots & \dots & \dots \\ \lambda_1 x_n^{(1)} & \lambda_2 x_n^{(2)} & \dots & \lambda_n x_n^{(n)} \end{pmatrix} \tag{49}$$

and it can be seen that we get the same result if we postmultiply X by the diagonal matrix Λ, whose diagonal elements are the eigenvalues:

$$X\Lambda = \begin{pmatrix} x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(n)} \\ x_2^{(1)} & x_2^{(2)} & \dots & x_2^{(n)} \\ \dots & \dots & \dots & \dots \\ x_n^{(1)} & x_n^{(2)} & \dots & x_n^{(n)} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} = \begin{pmatrix} \lambda_1 x_1^{(1)} & \lambda_2 x_1^{(2)} & \dots & \lambda_n x_1^{(n)} \\ \lambda_1 x_2^{(1)} & \lambda_2 x_2^{(2)} & \dots & \lambda_n x_2^{(n)} \\ \dots & \dots & \dots & \dots \\ \lambda_1 x_n^{(1)} & \lambda_2 x_n^{(2)} & \dots & \lambda_n x_n^{(n)} \end{pmatrix} \tag{50}$$

and we conclude that

$$\Lambda = X^{-1} A X \tag{51}$$

or, because of the unitarity of X

$$\Lambda = X^\dagger A X \tag{52}$$

This procedure of carrying out the transformation of A to the diagonal form Λ is called *diagonalization*. The matrix X is called the *diagonalising matrix*.

Important about the diagonalization is that Λ has the same eigenvalues as A. This can be seen immediately by writing down the characteristic equation of Λ and comparing with Eq. (45). Thus the expressions diagonalization and solving the eigenvalue problem are used interchangably.

9.12) Exercises.

9.12.1) Find the eigenvalues and the normalized eigenvectors of the symmetric matrix $A = \begin{pmatrix} 4 & -2 \\ -2 & 2 \end{pmatrix}$ and verify that the eigenvectors are orthogonal.

Answer: $3 \pm \sqrt{5}$; $\dfrac{1}{\sqrt{10 \mp 2\sqrt{5}}} \begin{pmatrix} 2 \\ 1 \mp \sqrt{5} \end{pmatrix}$.

9.12.2) Find the eigenvalues and the normalized eigenvectors of the symmetric matrix $A = \begin{pmatrix} 9 & \sqrt{3} \\ \sqrt{3} & 11 \end{pmatrix}$ and verify that the eigenvectors are orthogonal.

Answer: 2,3; $\begin{pmatrix} \sqrt{3}/2 \\ -1/2 \end{pmatrix}$, $\begin{pmatrix} 1/2 \\ \sqrt{3}/2 \end{pmatrix}$.

9.12.3) Find the eigenvalues and the normalized eigenvectors of the Hermitian matrix $A = \begin{pmatrix} 4 & i \\ -i & 2 \end{pmatrix}$ and verify that the eigenvectors are orthogonal.

Answer: $3 \pm \sqrt{2}$; $\frac{1}{2}\sqrt{2 \pm \sqrt{2}} \begin{pmatrix} 1 \\ i(1 \mp \sqrt{2}) \end{pmatrix}$

9.12.4) Find the eigenvalues and the normalized eigenvectors of the Hermitian matrix $A = \begin{pmatrix} \cos\alpha & i\sin\alpha \\ -i\sin\alpha & \cos\alpha \end{pmatrix}$ and verify that the eigenvectors are orthogonal. Write down the unitary diagonalising matrix U and verify by explicit multiplication that it gives the correct diagonalised form of A.

Answer: $\cos\alpha \pm \sin\alpha$; $\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \mp i \end{pmatrix}$; $U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}$.

9.12.5) Show that $A^2\vec{x} = \lambda^2\vec{x}$ if λ is an eigenvalue of A and $\vec{x}$ is the corresponding eigenvector.

9.12.6) Show that $A^{-1}\vec{x} = \lambda^{-1}\vec{x}$ if λ is an eigenvalue of A and $\vec{x}$ is the corresponding eigenvector.

9.12.7) Show that $p(A)\vec{x} = p(\lambda)\vec{x}$, where $p(A)$ is a polynomial of A, if λ is an eigenvalue of A and $\vec{x}$ is the corresponding eigenvector.

9.12.8) Show that A^T has the same eigenvalues and eigenvectors as A if A is a symmetric matrix.

9.12.9) Show that $A^\dagger$ has the same eigenvalues and eigenvectors as A if A is a Hermitian matrix.

9.13) Degenerate eigenvalues.

If the characteristic equation has a multiple root, then that eigenvalue is said to be *degenerate*. If the multiplicity of the root is k, then the eigenvalue is called k-fold degenerate.

It is always possible to find k mutually orthogonal eigenvectors to a k-fold degenerate eigenvalue.

Let us illustrate this statement with a simple example. Consider the symmetric matrix

$$A = \begin{pmatrix} 1 & 0 & -\sqrt{3} \\ 0 & 2 & 0 \\ -\sqrt{3} & 0 & -1 \end{pmatrix}$$

The characteristic equation of A is

$$\det(A - \lambda E) = -(\lambda - 2)^2(\lambda + 2) = 0$$

and hence the eigenvalues are $\lambda_1 = \lambda_2 = 2$ and $\lambda_3 = -2$, i.e. the eigenvalue $\lambda = 2$ is 2-fold degenerate. Substituting the degenerate eigenvalue into the eigenvalue equation we get

$$\begin{aligned} -x_1 - \sqrt{3}x_3 &= 0 \\ 0 \cdot x_2 &= 0 \\ -\sqrt{3}x_1 - 3x_3 &= 0 \end{aligned}$$

The first and last of these equations are identical and give

$$x_1 = -\sqrt{3}x_3$$

whereas the second equation admits any value of x_2. To get the two vectors that we need we can proceed like this:
(i) Set $x_3 = 0$ and $x_2 = 1$, hence

$$\vec{x}^{(1)} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

(ii) set $x_3 = 1$ and $x_2 = 0$, hence

$$\vec{x}^{(2)} = \frac{1}{2} \begin{pmatrix} -\sqrt{3} \\ 0 \\ 1 \end{pmatrix}$$

where we have already included the appropriate normalization factor. It can be readily checked that the two vectors are orthogonal.

The present example was so simple that we could see directly how to get the two orthogonal eigenvectors. But this may not always be the case. However, we can always construct normalized eigenvectors which are different from each other, and then apply a procedure to transform that set of nonorthogonal vectors into a set of orthogonal ones.

To illustrate this with the same example, we could also proceed like this:
(i) set $x_2 = 1$ and $x_3 = 1$ hence

$$\vec{x}^{(1)} = \frac{1}{\sqrt{5}} \begin{pmatrix} -\sqrt{3} \\ 1 \\ 1 \end{pmatrix} \tag{53}$$

(ii) set $x_2 = 1$ and $x_3 = -1$, hence

$$\vec{x}^{(2)} = \frac{1}{\sqrt{5}} \begin{pmatrix} \sqrt{3} \\ 1 \\ -1 \end{pmatrix} \tag{54}$$

These two vectors are both normalized eigenvectors of A with eigenvalue $\lambda = 2$, but they are not orthogonal: $\vec{x}^{(1)T}\vec{x}^{(2)} = -3/5$.

To see how we can transform this set of eigenvectors into a mutually orthogonal one, let us first establish the following:

> *any linear combination of eigenvectors $\vec{x}_i$ with the common eigenvalue λ is also an eigenvector of A*

Proof: Let $\vec{y} = \sum_{i=1}^{k} c_i \vec{x}(i)$ with arbitrary superposition coefficients c_i, hence

$$A\vec{y} = A \sum_{i=1}^{k} c_i \vec{x}^{(i)} = \sum_{i=1}^{k} c_i A \vec{x}^{(i)} = \sum_{i=1}^{k} c_i \lambda \vec{x}^{(i)} = \lambda \vec{y}$$

i.e. $\vec{y}$ is indeed eigenvector of A with eigenvalue λ.

Thus the strategy of transforming the set of nonorthogonal eigenvectors $\vec{x}^{(i)}$, $i = 1, 2, \ldots, k$, with eigenvalue λ into a set of orthonormal eigenvectors $\vec{y}^{(i)}$ consists of making a sequence of linear combinations of the $\vec{x}^{(i)}$, ensuring that they be orthogonal. This can be accomplished in the following way: set

$$\begin{aligned} \vec{y}^{(1)} &= \vec{x}^{(1)} \\ \vec{y}^{(2)} &= \vec{x}^{(2)} - (\vec{y}^{(1)\dagger}\vec{x}^{(2)})\vec{y}^{(1)} \end{aligned}$$

then $\vec{y}^{(1)}$ and $\vec{y}^{(2)}$ are orthogonal:

$$\vec{y}^{(1)\dagger}\vec{y}^{(2)} = (\vec{y}^{(1)\dagger}\vec{x}^{(2)}) - (\vec{y}^{(1)\dagger}\vec{x}^{(2)})(\vec{y}^{(1)\dagger}\vec{y}^{(1)}) = 0$$

where in the last step we have used the normalization of $\vec{y}^{(1)}$. If we now normalize $\vec{y}^{(2)}$ (without changing its notation), we can continue this iterative scheme:

$$\begin{aligned}
\vec{y}^{(3)} &= \vec{x}^{(3)} - (\vec{y}^{(2)\dagger}\vec{x}^{(3)})\vec{y}^{(2)} - (\vec{y}^{(1)\dagger}\vec{x}^{(3)})\vec{y}^{(1)} \\
\ldots & \quad \ldots \\
\vec{y}^{(k)} &= \vec{x}^{(k)} - (\vec{y}^{(k-1)\dagger}\vec{x}^{(k)})\vec{y}^{(k-1)} - \ldots - (\vec{y}^{(1)\dagger}\vec{x}^{(k)})\vec{y}^{(1)}
\end{aligned}$$

where it is understood that in each step we normalize the new vector $\vec{y}^{(i)}$.

Exercise: Carry out the orthogonalization of the eigenvectors (53) and (54).
Answer:

$$\vec{y}^{(1)} = \vec{x}^{(1)} = \frac{1}{\sqrt{5}}\begin{pmatrix} -\sqrt{3} \\ 1 \\ 1 \end{pmatrix}$$

$$\begin{aligned}
\vec{y}^{(2)} &= \vec{x}^{(2)} - (\vec{y}^{(1)T}\vec{x}^{(2)})\vec{y}^{(1)} = \frac{1}{\sqrt{5}}\begin{pmatrix} \sqrt{3} \\ 1 \\ -1 \end{pmatrix} - \frac{1}{5\sqrt{5}}(-\sqrt{3}\ 1\ 1)\begin{pmatrix} \sqrt{3} \\ 1 \\ -1 \end{pmatrix}\begin{pmatrix} -\sqrt{3} \\ 1 \\ 1 \end{pmatrix} \\
&= \frac{1}{\sqrt{5}}\begin{pmatrix} \sqrt{3} \\ 1 \\ -1 \end{pmatrix} + \frac{3}{5\sqrt{5}}\begin{pmatrix} -\sqrt{3} \\ 1 \\ 1 \end{pmatrix} = \frac{1}{5\sqrt{5}}\begin{pmatrix} 2\sqrt{3} \\ 8 \\ -2 \end{pmatrix}
\end{aligned}$$

To normalise $y^{(2)}$ we find the square of its norm:

$$y^{(2)T}y^{(2)} = (4/5)^2$$

hence, without changing the notation, we get the normalised vector $y^{(2)}$:

$$y^{(2)} = \frac{1}{2\sqrt{5}}\begin{pmatrix} \sqrt{3} \\ 4 \\ -1 \end{pmatrix}$$

and it remains to verify that $y^{(1)}$ and $y^{(2)}$ are orthogonal:

$$y^{(1)T}y^{(2)} = \frac{1}{10}(-\sqrt{3}\ 1\ 1)\begin{pmatrix} \sqrt{3} \\ 4 \\ -1 \end{pmatrix} = 0.$$

10. Linear simultaneous equations.

A linear simultaneous equation with two unknowns x and y is of the form

$$a_{11}x + a_{12}y = b_1 \tag{55}$$

$$a_{21}x + a_{22}y = b_2 \tag{56}$$

Each one of the two equations represents a straight line in the xy plane. Rewriting Eq. (55) in the form of

$$y = -\frac{a_{11}}{a_{12}}x + \frac{b_1}{a_{12}}$$

we see that $-a_{11}/a_{12}$ is the *slope* of the line and b_1/a_{12} is its *intercept*, i.e. the value of y at $x = 0$. Similarly $-a_{21}/a_{22}$ is the slope of line (56) and b_2/a_{22} is its intercept.

The two lines have either one point in common, if they cross, or they have no point in common, if they are parallel to each other, or they have all points in common, if they fall on top of each other.

To have one and only one point in common, the lines must have different slopes. In this case the simultaneous equations have one and only one solution, namely the crossing point of the lines. If the slopes are equal, then the lines are parallel if their intercepts are different and they fall on top of each other if their intercepts coincide. In the former case the simultaneous equations have no solutions, in the latter case they have infinitely many solutions, namely any pair of values x and y that satisfy either one of equations (55) or (56). Expressing these three cases in symbols we have:

$$\begin{array}{lll} \text{(i)} & a_{11}/a_{12} \neq a_{21}/a_{22}, & & \text{unique solution} \\ \text{(ii)} & a_{11}/a_{12} = a_{21}/a_{22}, & b_1/a_{12} \neq b_2/a_{22}, & \text{no solution exists} \\ \text{(iii)} & a_{11}/a_{12} = a_{21}/a_{22}, & b_1/a_{12} = b_2/a_{22}, & \text{infinitely many solutions} \end{array}$$

Of particular interest is the case of the existence of a unique solution. We can rewrite the corresponding condition (i) in the form of

$$a_{11}a_{22} - a_{21}a_{12} \neq 0$$

and we recognise that the l.h. side of this equation is the determinant of the coefficient matrix $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$. We can therefore write the condition for the existence of a unique solution also in the form of

$$\det A \neq 0 \tag{57}$$

The explicit form of the solution, which can be found for instance by elimination, is given by

$$x = \frac{1}{\Delta} \begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix}, \qquad y = \frac{1}{\Delta} \begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix} \tag{58}$$

where $\Delta = \det A$.

Considering simultaneous equations with more than two unknowns, it is convenient to denote them by a single letter with a subscript, numbering the unknowns from 1 to n, i.e. we write $x_1, x_2, \ldots, x_n$. The linear simultaneous equations take on the form of

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n & = & b_2 \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots & = & \ldots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n & = & b_n \end{array} \tag{59}$$

or, in matrix notation,

$$A\vec{x} = \vec{b} \tag{60}$$

where

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix}, \qquad \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{pmatrix}, \qquad \text{and} \qquad \vec{b} = \begin{pmatrix} b_1 \\ b_2 \\ \ldots \\ b_n \end{pmatrix}.$$

In direct generalization of the case of two equations the condition for the existence of a unique solution is again of the form of Eq. (57), and the solution itself is given by

$$x_i = \frac{\Delta_i}{\Delta} \tag{61}$$

where $\Delta = \det A$ and Δ_i is the determinant, obtained from $\det A$ by replacing the ith column by the vector $\vec{b}$.

Now, although Eq. (61) gives in principle the complete solution of the linear simultaneous equation, in practice it is more convenient to use the elimination method, also known as the Gaussian elimination method. For two or even for three unknowns the elimination method can be applied by hand. If there are more than three unknowns one will usually resort to a computer.