A. W. JOSHI

MATRICES AND TENSORS IN PHYSICS

MATRICES and TENSORS in Physics

A. W. JOSHI
Meerut University Institute of Advanced Studies

WILEY EASTERN LIMITED
New Delhi - Bangalore - Bombay

Reprinted 1977

Library of Congress Cataloging in Publication Data

Joshi, A W
Matrices and tensors in physics

Includes bibliographies.
1. Matrices. 2. Calculus of tensors.
3. Mathematical physics. I. Title.

QC20.7.M3J67 1975 530.1'22 75-26772

ISBN 0 85226 441 0 Paper
ISBN 0 85226 442 9 Cloth

Published by Vinod Kumar for Wiley Eastern Limited, AB 8 Safdarjang Enclave, New Delhi 110016 and printed by Pramodh Kapur at Raj Bandhu Industrial Company, C-61, Mayapuri, Phase-II, New Delhi 110 064.

Printed in India.

Preface

> **"Through understanding of what, pray, does all this world become understood, Sir?"**
>
> **Muṇḍaka Upaniṣad, 1-1-3**

The present century has seen the development of modern physics along two major directions—the theories of relativity and quantum mechanics. For a proper understanding of the basic concepts of quantum mechanics, a sound foundation in matrix algebra and vector spaces is essential. Similarly, tensor analysis and differential geometry provide the mathematical foundation for the theories of relativity. It is from this point of view that matrix algebra and tensor analysis have been selected for presentation in this book.

Matrices and tensors are generally taught to B.Sc. and M.Sc. (physics) students in Indian universities in a course on mathematics or mathematical methods in physics. Matrices are generally taken up directly without recourse to vector spaces and transformations. However, the connecting link between matrices and quantum mechanics is vector spaces, and it is often found that students fail to grasp the fundamental principles of quantum mechanics because of a lack of understanding of vector spaces. The first chapter of this book starts by introducing vector spaces and transformations, and matrices arising from it. This also gives the proper historical perspective because the study of matrices arose in connection with transformations in vector spaces. Emphasis is given to important concepts needed in quantum mechanics, such as linear combinations, linear dependence and independence of vectors and matrices, number of independent parameters of various special matrices, the most general matrix of a given type, etc. Section 9 contains an improved proof of the theorem that two matrices can be simultaneously diagonalized if and only if they commute with each other. Topics such as nondiagonalizable matrices and functions of a matrix are also included. Infinite (discrete and continuous) matrices are also discussed briefly at the end of chapter I.

The second chapter deals with the algebra and the calculus of general tensors in an N-dimensional Riemannian space. Many physical examples are given showing the necessity of the use of second and higher rank tensors. These include the conductivity tensor, the effective mass tensor, stress and strain tensors, tensors of elastic stiffness and compliance constants, piezoelectric strain coefficient tenser, dielectric susceptibility tensor, moment of inertia tensor, curvature tensor, etc. In this chapter, emphasis is given on the proper use of free and dummy indices in tensor equations. Einstein's summation convention is explained in detail and common errors arising in its use are pointed out. Rules are given to check the correctness of indices in a tensor equation.

Although it is desirable to have an elementary knowledge of vector spaces for a better understanding of the use of matrices and of the eigenvalue problem, the matter is so arranged that a reader wishing to proceed straight to matrices without going through vector spaces may skip Section 1 without much loss of continuity. However, it is felt that this is not the correct approach.

There are two appendices in the book. Appendix I discusses the logic of the phrases such as "if", "only if", and "if and only if". Appendix II gives a proof of the result that a finite nonsingular matrix (that is, one possessing both left and right inverses) must be square.

Altogether about 100 solved examples are included, and about 200 exercises are given in the book. Many of the exercises anticipate what is to come in the succeeding sections. Answers to selected exercises are given at the end of the book and such exercises are marked with asterisks.

The book has developed from courses given to graduate and postgraduate students during the past six years. Students' problems and queries have helped a lot in the preparation and presentation of the material. I am also thankful to my colleagues and senior students for their help during the preparation of the manuscript. Suggestions and comments from readers will be welcome.

A. W. JOSHI

June, 1975

Set theoretical notation

$\in$: belongs to, belonging to
$\ni$: such that
$\exists$: there exists
$\forall$: for every, for all
$\subset$: is contained in
$\supset$: contains
$\Rightarrow$: implies, only if
$\Leftarrow$: is implied by, if, follows from
$\Leftrightarrow$: implies and is impli d by, if and only if
$\not\Rightarrow$: does not imply
$\not\Leftarrow$: is not implied by, does not follow from

EXAMPLE: An expression such as

$$\exists\, e \in G \ni xe = ex = x \;\forall\; x \in G$$

means that "there exists an element e belonging to G such that $xe = ex = x$ for all x belonging to G".

Contents

I Matrix Algebra

Matrices occur in physics mainly in two ways: first, in the solution of systems of linear equations, and second, in the solution of eigenvalue problems in classical and quantum mechanics. Both these types of problems arise, in turn, from transformations of vectors in vector spaces and the operation of linear operators on vector spaces. The methods of matrix algebra are not only useful but essential in handling such problems. In this chapter, we shall discuss various operations with matrices and different situations in which they can be applied.

1. Vector spaces and transformations

We shall begin by discussing how matrices arise in connection with vector spaces. For the sake of completeness, it will be convenient to define a *group* and a *field* before coming to vector spaces.

Group

Consider the set $R=\{x: x=\text{a real number}\}$, the set of all real numbers. The elements of R, endowed with a law of addition, satisfy four properties: (a) for any two real numbers x and y belonging to R, their sum $x+y$ belongs to R; (b) there exists an element 0 in R, called *zero*, such that $x+0=0+x=x$ for every $x\in R$; (c) for every $x\in R$, there exists a unique element y in R such that $x+y=y+x=0$; (d) for any three elements x, y and z of R, the associative law is satisfied, that is, $x+(y+z)=(x+y)+z$.

Similarly, consider another set, $U(1)=\{z: |z|=1\}$, the set of all complex numbers of unit magnitude. The elements of $U(1)$, endowed with a law of multiplication, satisfy four properties: (a) for any two complex numbers of unit magnitude x and y, their product is also a complex numbers of unit magnitude and hence belongs to $U(1)$; (b) there exists an element $1+0i$, where $i=\sqrt{-1}$, or 1 for short, in $U(1)$ such that $x1=1x=x$ for every element x of $U(1)$; (c) for every $x\in U(1)$, there exists a unique element y in $U(1)$ such that $xy=yx=1$; (d) for any three elements x, y and z of $U(1)$, the associative law is satisfied, that is, $x(yz)=(xy)z$.

We notice that the four properties satisfied by the two sets are similar in nature. In fact, both the sets considered above are examples of a group.

In general, a *group* is defined as a set $G=\{x, y, z\ldots\}$ endowed with a binary law of composition for its elements such that it satisfies the four properties listed below by using the set theoretical notation:

(a) CLOSURE: $xy\in G\,\forall\, x, y\in G$;
(b) EXISTENCE OF IDENTITY: $\exists e\in G \ni xe=ex=x\,\forall\, x\in G$;
(c) EXISTENCE OF INVERSE: $\forall x\in G\,\exists y\in G \ni xy=yx=e$; (1)
(d) ASSOCIATIVE PROPERTY: $x(yz)=(xy)z\,\forall\, x, y, z\in G$.

These are known as the *group axioms*. The element e satisfying property (b) is known as the *identity element* of the group, and if $xy=yx=e$ [property (c)], y is known as the *inverse* of x and vice versa.

It is not necessary that the law of composition of group elements be commutative. In fact, in general, it is not, so that $xy \neq yx \forall x, y \in G$, where G is a group. If, in particular, $xy = yx \forall x, y \in G$, G is said to be an *abelian group*. Both the examples discussed above are abelian groups.

Field

A *field* $F = \{a, b, c, d, \ldots\}$ is a set of elements, endowed with two binary laws of composition for its elements, one denoted by $+$ called *addition*, and the other denoted by $\circ$ called *multiplication*, such that the following two conditions are satisfied:

(a) F is an abelian group under addition with the identity element denoted by 0 and called *zero*; and

(b) the set of the nonzero elements of F is an abelian group under multiplication with the identity element denoted by 1 and called *unity*.

0 is called the *additive identity* while 1 the *multiplicative identity* of the field.

Examples of a field are:

1. The set R of all real numbers with the additive identity 0 and the multiplicative identity 1.
2. The set C of all complex numbers with the additive identity $0+0i$ and the multiplicative identity $1+0i$.
3. The set $\{0, 1, 2, \ldots, p-1\}$ of p integers, where p is a prime number greater than 1, with the two binary operations of addition modulo p and multiplication modulo p; a finite field is called a *Galois field*.

The elements of a field are called *scalars*.

Vector space

A set $L = \{\mathbf{u}, \mathbf{v}, \mathbf{w}, \ldots\}$ is said to be a *vector space over a field* F if the following two conditions are satisfied:

(a) An operation of addition denoted by $+$ is defined in L such that L is an abelian group under addition. The identity element of this group will be denoted by $\mathbf{0}$.

(b) A scalar of the field F and an element of the set L can be combined by an operation called *scalar multiplication* to give an element of L such that for every $\mathbf{u}, \mathbf{v} \in L$ and $a, b \in F$, we have

$$a(\mathbf{u}+\mathbf{v})=a\mathbf{u}+a\mathbf{v}\in L;\quad (a+b)\,\mathbf{u}=a\mathbf{u}+b\mathbf{u}\in L,$$

$$a(b\mathbf{u})=(a\circ b)\,\mathbf{u},\quad 1\mathbf{u}=\mathbf{u},\, 0\mathbf{u}=\mathbf{0}. \tag{2}$$

The elements of a vector space are called *vectors.* Also note that 0 is an element (the additive identity) of F, whereas $\mathbf{0}$ is the *null vector* of L. The phrases *linear vector space* and *linear space* are also used for a vector space. Henceforth, we shall drop the multiplication symbol for scalars and write, for example, $a\circ b$ simply as ab.

The familiar three-dimensional space of position vectors is an example of a vector space over the field of real numbers. First, it is evident that the set of all position vectors is an abelian group. For, if $\mathbf{u}$ and $\mathbf{v}$ are any two vectors of this space, $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$ is also a vector of this space. The identity element is the null vector $\mathbf{0}$. The 'inverse' of a vector $\mathbf{u}$ is the vector $-\mathbf{u}$, because $\mathbf{u}+(-\mathbf{u})=\mathbf{0}$. Moreover, the law of vector addition is associative. Second, the position vectors satisfy the properties listed in Eqs. (2) for all a, b belonging to the field R of real numbers.

Linear independence of vectors

Two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ of a vector space are said to be *linearly dependent* if one is a multiple of the other, i.e., if $\mathbf{u}=c\mathbf{v}$, where c is some scalar. In other words, $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent if it is possible to find scalars a and b different from zero such that $a\mathbf{u}+b\mathbf{v}=\mathbf{0}$. Note that this is equivalent to saying that $\mathbf{u}$ is a multiple of $\mathbf{v}$ and vice versa.

Conversely, two vectors $\mathbf{u}$ and $\mathbf{v}$ are said to be *linearly independent* of each other if one is not a multiple of the other. In this case, it is impossible to satisfy the equation $a\mathbf{u}+b\mathbf{v}=\mathbf{0}$ except when $a=b=0$.

The concepts of linear dependence and independence of vectors can be extended to more than two vectors. Consider a set of n vectors $\mathbf{x}_1, \mathbf{x}_2,\ldots,\mathbf{x}_n$, none of which is a null vector. The vectors of this set are said to be *linearly dependent* if it is possible to find scalars $a_1, a_2,\ldots,a_n$, at least two of which are nonzero, such that

$$a_1\mathbf{x}_1+a_2\mathbf{x}_2+\ldots+a_n\mathbf{x}_n=\mathbf{0}. \tag{3}$$

Suppose the coefficient a_i is different from zero. Then dividing Eq. (3) by a_i, we have

$$\mathbf{x}_i=b_1\mathbf{x}_1+\ldots+b_{i-1}\mathbf{x}_{i-1}+b_{i+1}\mathbf{x}_{i+1}+\ldots+b_n\mathbf{x}_n, \tag{4}$$

with $b_j=-a_j/a_i$. Thus we can say that *the n vectors of a set are linearly dependent if at least one of them can be expressed as a linear combination of the remaining n−1 vectors.*

Conversely, the n vectors of a set are *linearly independent* if the only solution of Eq. (3) is $a_i=0$ for $1 \leqslant i \leqslant n$. In other words, the vectors of a set are linearly independent if it is impossible to construct the null vector from a linear combination of the vectors except when all the coefficients vanish.

As an example, in the space of three-dimensional position vectors, it is possible to find a set of three, but not more than three, linearly independent vectors. If we choose the three linearly independent vectors to be $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, any other vector of the space can be expressed as a linear combination of these three vectors in the form

$$\mathbf{u}=a_1\mathbf{x}_1+a_2\mathbf{x}_2+a_3\mathbf{x}_3. \tag{5}$$

The *dimension* of a space is defined as the maximum number of linearly independent vectors in the space. A set of n linearly independent vectors in an n-dimensional vector space is called a *basis* for the vector space. Clearly, the basis is not unique and we may choose the basis in an infinite number of ways.

Vector space of n-tuplets

The vector space of n-tuplets is the most general vector space in the sense that *every vector space is a set of n-tuplets.* An n-tuplet is an ordered set of n numbers, real or complex, such as $\mathbf{u}=(u_1, u_2, \ldots, u_n)$, where u_i are any numbers, all of the same type (that is, real or complex). If we consider the set of all such elements $(\mathbf{u}, \mathbf{v}, \ldots)$ obtained by giving all possible values to the components u_i, we have the vector space of n-tuplets; u_i are said to be the *components* of the vector $\mathbf{u}$. A vector all of whose components are real numbers is said to be a real vector. A vector space whose vectors are real is called a *real vector space.* We shall in general consider the vector space of complex n-tuplets.

A vector all of whose components are zero is called the *null vector*: $\mathbf{0}=(0, 0, \ldots, 0)$. Two vectors of a vector space are said to be equal to each other if, and only if, their respective components are equal to each other. Thus,

$$\mathbf{u}=\mathbf{v} \Leftrightarrow u_i=v_i \text{ for } 1 \leqslant i \leqslant n. \tag{6}$$

The addition of two vectors and the scalar multiplication are defined as follows:

$$\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}=(u_1+v_1, u_2+v_2, \ldots, u_n+v_n), \tag{7a}$$

$$a\mathbf{u}=\mathbf{u}a=(au_1, au_2, \ldots, au_n), \tag{7b}$$

both of which can be seen to be elements of the vector space.

Inner product space

A vector space L defined over a field F, where F is the field of real or complex numbers, is called an *inner product space* if with every pair of elements $\mathbf{u}, \mathbf{v} \in L$, there is associated a unique number belonging to the field F—denoted by $(\mathbf{u}, \mathbf{v})$ and called the *inner product* or the *scalar product* of $\mathbf{u}$ and $\mathbf{v}$—for which the following properties hold:

$$(\mathbf{u}, \mathbf{v}) = (\mathbf{v}, \mathbf{u})^*,$$
$$(a\mathbf{u}, b\mathbf{v}) = a^*b\,(\mathbf{u}, \mathbf{v}), \tag{8}$$
$$(\mathbf{w}, a\mathbf{u}+b\mathbf{v}) = a(\mathbf{w}, \mathbf{u}) + b(\mathbf{w}, \mathbf{v}),$$

where the asterisk denotes the complex conjugate.

The vector space of n-tuplets of real or complex numbers can be made an inner product space if we define the inner product of two vectors by

$$(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^{n} u_i^* v_i. \tag{9}$$

The ordinary three-dimensional space of position vectors is also an inner product space with the familiar rule for taking the scalar product of two vectors.

EXAMPLE 1: Determine whether the four vectors

$$\mathbf{u} = (1, 2, 3), \quad \mathbf{v} = (2, 0, -1),$$
$$\mathbf{w} = (1, -1, 1), \ \mathbf{x} = (2, 1, 0) \tag{10}$$

are linearly dependent or independent.

Solution: We solve the equation

$$a\mathbf{u} + b\mathbf{v} + c\mathbf{w} + d\mathbf{x} = \mathbf{0} \tag{11}$$

for the unknown coefficients a, b, c, d. Substituting for the vectors from Eqs. (10) and using the rules of vector addition and scalar multiplication given in Eqs. (7), Eq. (11) becomes

$$(a+2b+c+2d,\ 2a-c+d,\ 3a-b+c) = (0, 0, 0). \tag{12}$$

This gives us the three simultaneous equations

$$a+2b+c+2d=0,\ 2a-c+d=0,\ 3a-b+c=0, \tag{13}$$

which have the solution

$$b = 12a/5,\ c = -3a/5,\ d = -13a/5,\ a \text{ arbitrary}. \tag{14}$$

Thus Eq. (11) can be satisfied without all the coefficients being zero. We can write Eq. (11), using Eq. (14) and taking $a = 5$, as

$$5\mathbf{u} + 12\mathbf{v} - 3\mathbf{w} - 13\mathbf{x} = \mathbf{0}. \tag{15}$$

The given vectors are therefore linearly dependent.

EXAMPLE 2: It can be shown that we can find a maximum of n linearly independent vectors in the vector space of n-tuplets. We can conveniently choose these n independent vectors or 'coordinate axes' as

$$\begin{aligned} \mathbf{e}_1 &= (1, 0, 0, \ldots, 0), \\ &\vdots \\ \mathbf{e}_i &= (0, 0, \ldots, 1_i, \ldots, 0), \\ &\vdots \\ \mathbf{e}_n &= (0, 0, 0, \ldots, 1), \end{aligned} \tag{16}$$

where 1_i means that unity occurs in the i-th position. We shall now prove the following two results.

(a) *The n vectors $(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ are linearly independent.*

Solution: We consider the equation

$$a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + \ldots + a_n\mathbf{e}_n = \mathbf{0}. \tag{17}$$

Substituting for the vectors from Eqs. (16), we see that Eq. (17) becomes

$$(a_1, a_2, \ldots, a_n) = (0, 0, \ldots, 0). \tag{18}$$

The only solution of Eq. (17) is therefore $a_i = 0$ for $1 \leqslant i \leqslant n$, showing that the n vectors are linearly independent.

(b) *Any vector of the space* (that is, *any n-tuplet*) *can be uniquely expressed as a linear combination of the n vectors $\mathbf{e}_i$.*

Solution: Consider a vector $\mathbf{u} = (u_1, u_2, \ldots, u_n)$. Let us try to express it in the form

$$\mathbf{u} = b_1\mathbf{e}_1 + b_2\mathbf{e}_2 + \ldots + b_n\mathbf{e}_n. \tag{19}$$

This equation has the solution $b_i = u_i$ for $1 \leqslant i \leqslant n$, so that Eq. (19) becomes

$$\mathbf{u} = \sum_{i=1}^{n} u_i\mathbf{e}_i. \tag{20}$$

If we take the scalar product of a vector with itself, we find from Eq. (9) that

$$(\mathbf{u}, \mathbf{u}) = \sum_{i=1}^{n} |u_i|^2. \tag{21}$$

The positive square root of this quantity is defined as the *norm* of the vector $\mathbf{u}$ and is denoted by

$$\|\mathbf{u}\| = (\mathbf{u}, \mathbf{u})^{1/2} = \left[\sum_{i=1}^{n} |u_i|^2\right]^{1/2}. \tag{22}$$

It is easy to recognize that this is the length of a vector in familiar language. If the norm of a vector is unity, it is called a *unit vector* or a *normalized vector*. It can be seen that all the vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ defined in Eqs. (16) are unit vectors.

If the scalar product of two vectors is zero, the vectors are said to be *orthogonal*. Thus two vectors **u** and **v** are orthogonal if, and only if,

$$(\mathbf{u}, \mathbf{v})=0. \tag{23}$$

Once again it can be verified that any two vectors of the set $\{\mathbf{e}_i\}$ are orthogonal. Combining this with the earlier result, we have

$$(\mathbf{e}_i, \mathbf{e}_j)=\delta_{ij}. \tag{24}$$

A set of vectors each of which is orthogonal to all the remaining vectors of the set is called an *orthogonal set*. If each of the vectors is further normalized to unity, it is called an *orthonormal set*. A vector can be normalized by dividing it by its norm. Thus, if **u** is any vector, $\mathbf{u}/\|\mathbf{u}\|$ is the normalized vector parallel to **u**.

Schmidt's orthogonalization method

It is possible to obtain a set of orthogonal, or in fact orthonormal, vectors starting from a set of nonorthogonal but linearly independent vectors. This can be done by a procedure known as the *Schmidt's orthogonalization method* which is discussed below.

Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ be a set of linearly independent vectors which are not necessarily orthogonal. It is required to obtain a set of orthogonal vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ starting from the original set of vectors. We proceed along the following steps:

1. Take $\mathbf{v}_1=\mathbf{u}_1$.
2. Let

$$\mathbf{v}_2=\mathbf{u}_2+a_{21}\mathbf{v}_1, \tag{25}$$

where a_{21} is a constant to be determined from the condition that $\mathbf{v}_2$ be orthogonal to $\mathbf{v}_1$, i.e., $(\mathbf{v}_1, \mathbf{v}_2)=0$. Taking the scalar product of $\mathbf{v}_1$ with $\mathbf{v}_2$ of Eq. (25) and equating it to zero, we get

$$(\mathbf{v}_1, \mathbf{u}_2)+a_{21}(\mathbf{v}_1, \mathbf{v}_1)=0 \Rightarrow a_{21}=-(\mathbf{v}_1, \mathbf{u}_2)/(\mathbf{v}_1, \mathbf{v}_1). \tag{26}$$

Thus we have two orthogonal vectors, $\mathbf{v}_1$ and $\mathbf{v}_2$.

3. Let

$$\mathbf{v}_3=\mathbf{u}_3+a_{32}\mathbf{v}_2+a_{31}\mathbf{v}_1, \tag{27}$$

where a_{31} and a_{32} are constants to be determined from the conditions that $\mathbf{v}_3$ be orthogonal to $\mathbf{v}_1$ and $\mathbf{v}_2$. This gives

$$(\mathbf{v}_1, \mathbf{v}_3)\equiv 0=(\mathbf{v}_1, \mathbf{u}_3)+a_{31}(\mathbf{v}_1, \mathbf{v}_1)\Rightarrow a_{31}=-(\mathbf{v}_1, \mathbf{u}_3)/(\mathbf{v}_1, \mathbf{v}_1);$$
$$(\mathbf{v}_2, \mathbf{v}_3)\equiv 0=(\mathbf{v}_2, \mathbf{u}_3)+a_{32}(\mathbf{v}_2, \mathbf{v}_2)\Rightarrow a_{32}=-(\mathbf{v}_2, \mathbf{u}_3)/(\mathbf{v}_2, \mathbf{v}_2). \tag{28}$$

Now we have three mutually orthogonal vectors, $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$.