INTRODUCTION to MATRIX METHODS in OPTICS

A. Gerrard and J. M. Burch

Introduction to Matrix Methods in Optics

A. Gerrard

University of Bath

J. M. Burch

National Physical Laboratory, Teddington

A Wiley-Interscience Publication

JOHN WILEY & SONS

London • New York • Sydney • Toronto

Library of Congress Cataloging in Publication Data:

Gerrard, Anthony.
Introduction to matrix methods in optics.

"A Wiley-Interscience publication"
1. Optics. 2. Matrices. I. Burch, James M., joint author. II. Title.

QC355.2.G47 535 72-21192
ISBN 0 471 29685 6

Printed in Great Britain by Unwin Brothers Limited, The Gresham Press, Old Woking, Surrey.

Preface

Our purpose in writing this book has been not to present new results but to encourage the adoption of simple matrix methods in the teaching of optics at the undergraduate and technical college level. Many of these methods have been known for some time but have not found general acceptance; we believe that the time has now come for lecturers to reconsider their value. We believe this partly because the use of matrices is now being taught quite widely in schools, and many students will already have glimpsed something of the economy and elegance with which, for a linear system, a whole wealth of input-output relations can be expressed by a single matrix.

A second reason is that, for more than a decade, the field of optics has been enriched enormously by contributions from other disciplines such as microwave physics and electrical engineering. Although an engineering student may be a newcomer to optics, he may well have encountered matrix methods during his lectures on electrical filters or transmission lines; we think he will welcome an optics course which, instead of barricading itself behind its own time-honoured concepts, links itself recognizably to other disciplines.

Another barrier which we believe matrix methods may help to blur is the classical separation of optics into compartments labelled 'geometrical' and 'physical'. The optics of today transcends all boundaries, and the student may share our delight when a ray-transfer matrix, based on purely geometrical considerations, predicts with almost perfect accuracy the diffraction behaviour of a Gaussian beam as it is generated in a laser resonator or propagated through an external system. Surely the spirit of Gauss particularly must rejoice at this versatility!

Hoping as we do that matrix methods may help to forge links between the various branches of optics and other subjects, we have sought to avoid any inconsistencies of nomenclature. Out of the several types of ray-transfer matrix that have been proposed, we have followed Sinclair (of The Institute of Optics, Rochester, N.Y.) in choosing a form which is always unimodular, and which is compatible with the (*ynv*) method of calculation and with modern work on laser resonators. In contrast with Halbach's nomenclature, these matrices are defined so that they tell us what output we shall obtain for a given input; this is the choice favoured by most workers, and it ensures consistency with the rest of the book in which we describe the established Jones and Mueller calculus for polarization problems. The student will, of course, encounter situations, for example in laser beam propagation, where he has to work backwards and find what input is needed to produce a given desired output. In nearly all such cases, however, he will be dealing with 2×2 unimodular matrices, inversion of which he will learn to tackle with relish.

We shall now discuss some of the limitations and omissions in what we have written, and then describe briefly the arrangement of the chapters.

Because this is an introductory text which assumes very little prior knowledge, we have confined our attention to just two topics - namely paraxial imaging and polarization. The first topic has the advantage that the concepts required initially are almost intuitive; the second serves to emphasize the transverse nature of light waves but does not demand a knowledge of electromagnetic theory. We should have liked to include a chapter on reflection and transmission of light by thin films and stratified media, but to do this properly we should have had to proceed via a derivation from Maxwell's equations of the coupled behaviour of the transverse electric and magnetic field components.

It would have been possible to give a more superficial treatment, in which the illumination is restricted to be at normal incidence and a coefficient of reflection is assumed for each individual surface, but we feel it is better to omit the subject altogether and refer the student who is interested to existing coverage in the literature. Other topics which we rejected, but which might have been suitable for a more advanced book, are Wolf's coherency matrix and the use of 3×3 or 4×4 matrices to describe reflection from a series of variously oriented mirror surfaces, for example in a reflecting prism.

Our first chapter is intended for those who have no previous acquaintance with matrix algebra. Using numerous worked examples, it introduces the basic ideas of rectangular matrix arrays and gives the rules for adding them and for forming matrix products. The section on square matrices concentrates for simplicity on the 2×2 matrix. After the transpose matrix and the determinant have been introduced, the problem of matrix inversion is discussed. This leads into a brief treatment of diagonalization, and we conclude by showing how the Nth power of a matrix can be determined (without memorizing Sylvester's theorem).

Chapter II is devoted to the paraxial imaging properties of a centred optical system. Defining a ray in terms of its height and its optical direction-cosine, we show how a ray-transfer matrix can be used to describe the change that occurs in these two quantities as the ray traverses a system. The two basic types of matrix that represent the effect of a simple gap or of a single refracting surface are combined to form the equivalent matrix of a thin lens, a thick lens or a complete optical system. It is shown how the properties of a system can be inferred from knowledge of its matrix, and conversely how the matrix elements can be determined experimentally. At the end of the chapter we extend the ray-transfer matrix to include reflecting as well as refracting elements. The text is again supported by worked examples, and an appendix shows how the aperture properties of a system can be determined.

In the first part of chapter III we review and tabulate the results so far obtained and use them to describe the radius of curvature of a wavefront, the optical length of a ray path and the étendue of a beam. We then consider optical resonators and show how a round trip between the mirrors of a resonator can be represented by a single equivalent matrix. In order to consider the effect of repeated traversal of the resonator, we now diagonalize its matrix and find that, for the so-called 'unstable' case, both the eigenvalues and the eigenvectors are real; the former represent the loss per transit due to 'beam walk-off' and the latter represent solutions for the radius of curvature of a self-perpetuating wavefront.

For the case of a 'stable' laser resonator, both eigenvalues and eigenvectors are complex; the former represent the phase shift per transit and the latter can be interpreted in terms of Kogelnik's complex curvature parameter to predict not only the divergence but also the spot width of the Gaussian beam that the laser will generate. Furthermore, if we have a mode-matching problem in which we must calculate the diffraction of a laser beam as it is propagated externally, this too can be solved very easily by using the ray-transfer matrix. We conclude by indicating the extension of these methods to distributed lens-like media. The use of an augmented matrix to handle residual misalignments is discussed in an appendix.

In chapter IV we consider two alternative matrix methods for handling problems in polarization. After reviewing the different kinds of polarized light, we introduce first the Stokes parameters and the 4×4 Mueller matrices by which problems involving both polarized and unpolarized light can be tackled. The discussion includes numerous worked examples and an account of how both the Stokes parameters and the elements of a Mueller matrix can be determined experimentally. The Mueller matrices that are likely to be needed are tabulated and their derivation is given in an appendix.

A similar discussion is then given for the Jones calculus, which uses 2×2 complex matrices and is more suitable for dealing with fully polarized light. The material is so arranged that the student can, if he wishes, concentrate exclusively either on the Jones or on the Mueller method.

Other appendixes to this chapter contain a statistical treatment of the Stokes parameters and a full analysis of the connection between the elements of a Jones matrix and those of the corresponding Mueller matrix.

Chapter V is concerned with the application of matrix methods to the propagation of light in uniaxial crystals. Although it is more advanced and demands some knowledge of electromagnetic theory, it does not depend on the contents of chapter IV and can be read separately. We hope that the student who reads it will go on to some of the topics that we have omitted. A bibliography is provided.

Finally, since the chapters of this book have been designed primarily as educational stepping-stones and as illustrations of the matrix approach, only a limited range of optics has been covered. How far is a student likely to find this material of value during his subsequent career?

For the small fraction of graduates who will spend their life in optical design or manufacture, it has to be admitted that problems involving polarization do not often arise, and in most cases the contribution of first-order optics is trivial; the real problems arise either in using a computer to control third-order and higher-order aberrations or in more practical aspects of fabrication and assembly.

But for every professional optician there will be many others whose practice it will be to buy their optical equipment off the shelf and then incorporate it into larger systems. Some of these workers will be found in scientific research and others in new industries based on optoelectronics and laser engineering, but many will be engaged in more traditional fields such as mechanical engineering. We refer here not only to photoelasticity and to established techniques for optical inspection and alignment but also to more recent developments in laser holography and speckle interferometry. The era has already arrived where optical methods of measurement can be applied to a wide variety of unconventional tasks, and the engineer concerned usually has to work on a 'do it yourself' basis.

In holographic non-destructive testing, in the study of vibrations or in strain analysis, the component being viewed may be so irregular in shape that there is no point in trying to achieve well-corrected imaging over a wide flat field. It follows that money (as well as internal reflections) can often be saved by using the simplest of lenses; but before those cheap lenses are thrown together, with or without the help of a ray-transfer matrix, we hope that the reader of this book will at least remember to test each of them in a strain-viewer!

One of us, J. M. Burch, wishes to acknowledge the support of the Science and Technology Foundation of New York State which enabled him to spend a year as visiting professor at The Institute of Optics, University of Rochester. He is grateful to several of his Rochester colleagues, and notably to Douglas C. Sinclair, for discussions on some aspects of ray-transfer matrices.

Contents

I

Introduction to Matrix Calculations

I.1 INTRODUCTORY DISCUSSION

In this book we consider how some simple ideas of matrix algebra can be applied with advantage to problems involving optical imaging and polarization. The discussion in this chapter is designed mainly for those readers who have not so far encountered matrices or determinants; the treatment is elementary and covers only what will be needed to understand the rest of the book.

Matrices were introduced in 1857 by the mathematician Cayley as a convenient shorthand notation for writing down a whole array of linear simultaneous equations. The rules for operating with matrix arrays are slightly different from those for ordinary numbers, but they were soon discovered and developed. Matrix methods became of great interest to the physicist in the 1920's when Heisenberg introduced the matrix form of quantum mechanics. They are used in many kinds of engineering calculation but their application to optics is more recent.

Determinants, with which we shall be concerned to a lesser extent, were introduced by Vandermonde as early as 1771. They were at first called 'eliminants' because they arose in solving equations by the method of successive elimination. In most of the optical problems with which we shall deal the determinants all have a value of unity, and this fact provides a convenient check at the end of a calculation.

Let us now consider how the notion of a matrix arises. Suppose we have a pair of linear equations

$$U = Ax + By$$
$$V = Cx + Dy$$

where A, B, C and D are known constants, and x and y are variables. These equations enable us to calculate U and V if x and y are known. It proves convenient, for many purposes, to separate the constants from the variables. We write the pair of equations thus:

$$\begin{bmatrix} U \\ V \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

a *single* equation which is defined as meaning exactly the same as the pair. We regard each of the groups of symbols enclosed between a pair of vertical brackets as a single entity, called a MATRIX. $\begin{bmatrix} U \\ V \end{bmatrix}$ and $\begin{bmatrix} x \\ y \end{bmatrix}$ are called 'column matrices' or, alternatively, 'column vectors', since each contains only a single column.

The general matrix is a rectangular array with the symbols arranged in rows and columns. The matrix $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$, which has two rows and columns, is called a 'square matrix of order two'. Later we shall meet 'row matrices' (sometimes called 'row vectors') like $\begin{bmatrix} P & Q \end{bmatrix}$, in which the separate symbols, called 'matrix elements', are written horizontally in a single row. A matrix with only one element is just an ordinary number, or scalar quantity.

If we use a single symbol for each matrix, we can write the pair of equations even more briefly, thus:

$$C_2 = SC_1$$

where C_1 denotes the column matrix $\begin{bmatrix} x \\ y \end{bmatrix}$, C_2 denotes the column matrix $\begin{bmatrix} U \\ V \end{bmatrix}$ and S denotes the square matrix $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$.

Now let us suppose that U and V are linked in turn with another pair of variables, L and M, say, by another pair of linear equations, thus:

$$L = PU + QV$$
$$M = RU + TV$$

which we write in the form

$$\begin{bmatrix} L \\ M \end{bmatrix} = \begin{bmatrix} P & Q \\ R & T \end{bmatrix} \begin{bmatrix} U \\ V \end{bmatrix}$$

that is

$$C_3 = KC_2$$

where C_3 denotes $\begin{bmatrix} L \\ M \end{bmatrix}$ and K denotes $\begin{bmatrix} P & Q \\ R & T \end{bmatrix}$. We can, of course, find L and M in terms of x and y by substituting for U and V in the equations defining L and M. Thus:

$$L = P(Ax + By) + Q(Cx + Dy)$$

$$M = R(Ax + By) + T(Cx + Dy)$$

that is

$$L = (PA + QC)x + (PB + QD)y$$

$$M = (RA + TC)x + (RB + TD)y$$

which we write as

$$\begin{bmatrix} L \\ M \end{bmatrix} = \begin{bmatrix} PA + QC & PB + QD \\ RA + TC & RB + TD \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

that is

$$C_3 = FC_1$$

where F denotes $\begin{bmatrix} PA + QC & PB + QD \\ RA + TC & RB + TD \end{bmatrix}$. But, on the other hand, we can write

$$C_3 = KC_2 = K(SC_1)$$

Now, if this were an equation in ordinary algebra, we could rewrite it as

$$C_3 = KSC_1 = (KS)C_1$$

merely changing the positions of the brackets. KS would be called the product of K and S.

Again, comparing the equations linking C_1 and C_3, we could write

$$C_3 = KSC_1 \quad \text{and} \quad C_3 = FC_1$$

Therefore

$$F = KS$$

and we would say that F was the *product* of K and S.

In matrices we wish to follow a similar method but we now need to *define* the product of two matrices, since only products of single numbers are defined in ordinary algebra.

I.2 MATRIX MULTIPLICATION

We *define* matrix multiplication so that the above formalism can be carried over from ordinary algebra to matrix algebra. Thus, we *define* the product of the matrices by stating that K multiplied by S gives the product matrix F; that is

$$\begin{bmatrix} P & Q \\ R & T \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} PA + QC & PB + QD \\ RA + TC & RB + TD \end{bmatrix}$$

Examining the structure of the right-hand matrix (the product), it is easy to see how it is formed.

The top left-hand element is in the first row and the first column. It is produced by taking the first *row* of K, which is $\begin{bmatrix} P & Q \end{bmatrix}$, and the first *column* of S, which is $\begin{bmatrix} A \\ C \end{bmatrix}$, multiplying corresponding elements together (the first element of the row by the first element of the column, the second element of the row by the second element of the column), forming the products PA and QC, and then adding to get $PA + QC$.

The element in the first row and the second column of F is formed in the same way from the first row of K and the second column of S. The element in the second row and first column of F is formed from the second row of K and the first column of S. Finally, the element in the second row and second column of F is formed from the second row of K and the second column of S.

It proves useful, in some applications, to use a suffix notation for the elements of the matrices. We write a column matrix A, for instance, as

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$$

the subscript indicating the position of the element in the column. A square matrix S we write as

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$$

where the first subscript indicates which row an element is in and the second subscript indicates which column. If we re-express our two square matrices K and S in this suffix notation

$$K = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} \quad \text{and} \quad S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$$

then the product $F = KS$ becomes

$$\begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} = \begin{bmatrix} K_{11}S_{11} + K_{12}S_{21} & K_{11}S_{12} + K_{12}S_{22} \\ K_{21}S_{11} + K_{22}S_{21} & K_{21}S_{12} + K_{22}S_{22} \end{bmatrix}$$

that is

$$\begin{bmatrix} \sum_{i=1}^{2} K_{1i}S_{i1} & \sum_{i=1}^{2} K_{1i}S_{i2} \\ \sum_{i=1}^{2} K_{2i}S_{i1} & \sum_{i=1}^{2} K_{2i}S_{i2} \end{bmatrix}$$

This suggests a general formula for any element of the matrix:

$$F_{RT} = \sum_{i=1}^{i_{max}} K_{Ri}S_{iT}$$

where F_{RT} denotes the element in the Rth row and the Tth column of F, and similarly for K and S. (The summation sign used here indicates that the repeated suffix i takes on all possible values in succession; it is sometimes omitted.)

So far, we have confined our attention to two-by-two matrices and two-by-one columns; but the matrix idea is much more general than this. In this book we shall need two-by-two, three-by-three and four-by-four square matrices, two-by-one, three-by-one and four-by-one columns, and one-by-two, one-by-three and one-