Matrix Theory

DAVID W. LEWIS

World Scientific

Matrix Theory

DAVID W. LEWIS
Department of Mathematics
University College Dublin
Ireland

World Scientific
Singapore • New Jersey • London • Hong kong

Published by

World Scientific Publishing Co. Pte. Ltd.
P O Box 128, Farrer Road, Singapore 9128
USA office: Suite 1B, 1060 Main St., River Edge, NJ 07661
UK office: 73 Lynton Mead, Totteridge, London N20 8DH

Library of Congress Cataloging-in-Publication data is available.

MATRIX THEORY

ISBN 981-02-0689-5

Printed in Singapore by Utopia Press.

Matrix Theory

PREFACE

This book provides an introduction to matrix theory, and aims to provide a clear and concise exposition of the basic ideas, results and techniques in the subject. It combines the algebraic and analytic aspects of matrix theory. It presumes no knowledge beyond school mathematics although some familiarity with elementary calculus would be helpful in a few of the applications. It is hoped that the book can be profitably used by a wide range of students, including students of mathematics, engineering, science, and other disciplines where matrices arise. Complete proofs are given, although some are relegated to appendices at the end of chapters. This should enable the book to be used both by students who want all of the theory and those who are mainly interested in learning the techniques. The text is interspersed with many examples, applications and numerous exercises for the reader. Students who have already done an introductory linear algebra course may use the later chapters for a more advanced course.

Acknowledgements

I am indebted to my colleagues Fergus Gaines, Rod Gow and Tom Laffey for many useful discussions on matrices and linear algebra, and for reading the text and correcting some errors. I am grateful to my wife Anne for her constant support and encouragement during the preparation of this book.

TABLE OF CONTENTS

Chapter 1 Matrices and linear equations 1

1.1 Matrices and matrix algebra 1

1.2 Systems of linear equations 7

1.3 The inverse of a square matrix 19

1.4 Determinants 23

1.5 Some special types of matrices 29

1.6 More on systems of linear equations 30

1.7 Some places where matrices are found 33

Appendix: Proofs of the properties of determinants 38

Chapter 2 Vector spaces and linear maps 45

2.1 Vector spaces 46

2.2 Bases and dimension 51

2.3 Linear maps and isomorphisms 57

2.4 Subspaces and quotient spaces 62

2.5 Matrices and linear maps 67

2.6 The image and kernel of a linear map 71

2.7 Inner products 78

2.8 Miscellaneous comments 88

Chapter 3 Matrix norms 93

3.1 Norms on vector spaces 94

3.2 Analytic concepts 96

3.3 Norms on spaces of matrices 102

3.4 Examples and properties of operator norms 106

3.5 Matrix functions defined by power series 108

3.6 A criterion for invertibility 110

3.7 Stability of algorithms 112

Appendix: The equivalence of all norms in $\mathbb{R}^n$ and in $\mathbb{C}^n$ 113

Chapter 4 Eigenvalues and eigenvectors 116

4.1 Eigenvalues and eigenvectors 116

4.2 Determination of eigenvalues 117

4.3 Some properties of eigenvalues 120

4.4 Eigenvalues and norms 126

4.5 Polynomials in a matrix 127

4.6 Diagonalizable matrices 133

4.7 Some places where eigenvalues and eigenvectors are used 137

Chapter 5 Jordan canonical form and applications 141

5.1 Nilpotent matrices 141

5.2 Jordan canonical form 143

5.3 Further remarks on the Jordan canonical form 150

5.4 Systems of linear differential equations 155

5.5 Markov processes 163

5.6 Sylvester matrix equation 170

Appendix: Proof of the Jordan form theorem 178

Chapter 6 Symmetric and hermitian matrices 182

6.1 Schur's unitary triangularization theorem 183

6.2 Real quadratic forms 189

6.3 Hermitian forms 194

6.4 Reduction to diagonal form 198

6.5 Positive-definite forms and matrices 199

6.6 The signature of a quadratic or hermitian form 204

6.7 Simultaneous reduction of a pair of forms 211

6.8 Eigenvalues and the Rayleigh quotient 215

6.9 Courant-Fisher min-max theorem 222

Chapter 7 Perturbation theory 229

7.1 Gershgorin's theorem and Taussky's theorem 229

7.2 The condition of eigenvalues 236

7.3 Approximation by diagonalizable matrices 240

7.4 Condition numbers and their applications 243

7.5 Inverses and perturbed matrices 247

Chapter 8 Further topics 252

8.1 Normal matrices and their eigensystems 253

8.2 The QR-factorization and the QR-algorithm 256

8.3 The singular value decomposition 260

8.4 Generalized inverses of matrices 272

8.5 Least squares problems 276

8.6 Gauss-Seidel iteration 282

References and further reading 290

Answers to selected problems 291

Index 295

Chapter 1

MATRICES AND LINEAR EQUATIONS

A familiarity with matrices is necessary nowadays in many areas of mathematics and in a wide variety of other disciplines. Areas of mathematics where matrices occur include algebra, differential equations, calculus of several variables, probability and statistics, optimization, and graph theory. Other disciplines using matrix theory include engineering, physical sciences, biological sciences, economics and management science.

In this first chapter we give the fundamentals of matrix algebra, determinants, and systems of linear equations. At the end of the chapter we give some examples of situations in mathematics and other disciplines where matrices arise.

1.1 Matrices and matrix algebra

A *matrix* is a rectangular array of symbols. In this book the symbols will usually be either real or complex numbers. The separate elements of the array are known as the *entries* of the matrix.

Let m and n be positive integers. An $m \times n$ matrix A consists of m rows and n columns of numbers written in the following manner.

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

We often write $A = (a_{ij})$ for short. The entry a_{ij} lies in the i-th row and the j-th column of the matrix A.

Two $m \times n$ matrices $A = (a_{ij})$ and $B = (b_{ij})$ are *equal* if and only if all the corresponding entries of A and B are equal.

i.e. $a_{ij} = b_{ij}$ for each i and j.

The *sum* of the $m \times n$ matrices $A = (a_{ij})$ and $B = (b_{ij})$ is the $m \times n$ matrix denoted A + B which has entry $a_{ij} + b_{ij}$ in the (i,j)-place for each i,j.

Let λ be a scalar (i.e. a real or complex number) and let $A = (a_{ij})$ be an $m \times n$ matrix. The *scalar multiple* of A by λ is the $m \times n$ matrix denoted λA which has entry λa_{ij} in the (i,j)-place for each i,j.

1.1.2 **Proposition**

The following properties hold.

(i) $A + B = B + A$ for all $m \times n$ matrices A and B, i.e. addition of matrices is commutative.

(ii) $(A + B) + C = A + (B + C)$ for all $m \times n$ matrices A,B, and C, i.e.addition of matrices is associative.

(iii) $\lambda(A + B) = \lambda A + \lambda B$ for all scalars λ and all $m \times n$ matrices A and B.

(iv) $(\lambda_1 + \lambda_2)A = \lambda_1 A + \lambda_2 A$ for all scalars λ_1, λ_2 and all $m \times n$ matrices A.

(v) $(\lambda_1 \lambda_2)A = \lambda_1(\lambda_2 A)$ for all scalars λ_1, λ_2 and all mxn matrices A.

Proof

These properties follow at once from the properties of the real and complex number systems.

1.1.3 **Remark**

If we write -A for the matrix whose entries are $-a_{ij}$ for each i,j then $-A = (-1)A$, i.e. the multiple of the matrix A by the scalar -1. Also if we denote by O the nxn matrix with zero as each entry then $A + (-A) = O$.

1.1.4 Matrix multiplication

A 1xn matrix will be called a *row vector of length n* and an mx1 matrix will be called a *column vector of length m.*

Let $A = (a_1\ a_2\ a_3\ .\ .\ .\ a_n)$ be a 1xn matrix and $B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$ be an nx1 matrix.

We define the *product* AB to be the 1x1 matrix with the single entry

$$a_1b_1 + a_2b_2 + \ldots\ldots + a_nb_n.$$

Now we will define matrix multiplication in general. We say that the product AB of the two matrices A and B is defined if and only if the number of columns of A equals the number of rows of B.

(i.e. AB is defined if and only if A is an mxn matrix and B is an nxp matrix for some integers m,n,p.)

We define the *matrix product* AB to be the mxp matrix which has as its (i,j)-entry

$$a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \ldots\ldots + a_{in}b_{nj}$$

(In shorthand notation the (i,j)-entry is $\sum_{k=1}^{n} a_{ik}b_{kj}$.)

In other words the (i,j)-entry of AB is the product of the i-th row of A with the j-th column of B, this product being as in the special case of 1xn and nx1 matrices defined above.

1.1.5 Example

Let $A = \begin{pmatrix} 1 & -1 & 6 \\ 4 & 1 & -2 \\ 3 & 2 & 0 \end{pmatrix}$, $B = \begin{pmatrix} -1 & 6 & 1 & 2 \\ 0 & -2 & 2 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$.

Then $AB = \begin{pmatrix} 1 & -1 & 6 \\ 4 & 1 & -2 \\ 3 & 2 & 0 \end{pmatrix}\begin{pmatrix} -1 & 6 & 1 & 2 \\ 0 & -2 & 2 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 14 & 5 & 7 \\ -6 & 20 & 4 & 7 \\ -3 & 14 & 7 & 8 \end{pmatrix}$.

1.1.6 **Proposition**

The following properties hold.

(i) (AB)C = A(BC) whenever these products are meaningful.

(i.e. matrix multiplication is associative).

(ii) A(B + C) = AB + AC for all m×n matrices A,B and all n×p matrices C.

(iii) (A +B)C = AC + BC for all m×n matrices A and B and all n×p matrices C.

Proof

(i) Let A,B,C be of sizes m×n, n×p, p×q respectively.

The (i,k)-entry of AB is $\sum_{r=1}^{n} a_{ir} b_{rk}$ and hence the (i,j)-entry of (AB)C is $\sum_{r=1}^{n}\sum_{k=1}^{p} a_{ir} b_{rk} c_{kj}$. An examination of the product A(BC) shows that exactly the same expression occurs as the (i,j)-entry of A(BC).

(ii) Let A be of size m×n, B and C of size n×p.

Then the (i,j)-entry of A(B + C) is $\sum_{k=1}^{n} a_{ik}(b_{kj} + c_{kj})$ and this is easily seen to equal the (i,j)-entry of AB + AC.

(iii) This follows in a similar manner to (ii).

1.1.7 **Remark**

Matrix multiplication is not in general commutative. Note first that for AB and BA to both be defined it is necessary that A and B are each n×n matrices for some integer n, i.e. square matrices of the same size. However AB and BA will be different in general.

1.1.8 **Exercise**

Let $A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$, $B = \begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix}$. Show that AB ≠ BA.

1.1.9 **Remark**

A matrix of especial importance is the n×n *identity matrix*, denoted I_n, which is defined to have entries $a_{ii} = 1$ for all i and $a_{ij} = 0$ for i ≠ j. Often when we are dealing with n×n matrices for a fixed value of n we will simply write I for the identity matrix omitting the suffix n.

For any m×n matrix A it is easy to see that $AI_n = A$ and that $I_m A = A$.

1.1.10 **The transpose of a matrix**

Let A be an m×n matrix.

The *transpose* of A is the n×m matrix with entry a_{ji} in the (i,j)-place. The transpose of A is denoted by A^t.

Note that the rows of A become the columns of A^t and vice versa.

1.1.11 **Proposition**

The transpose satisfies the following properties.

(i) $(A + B)^t = A^t + B^t$ for all m×n matrices A and B.

(ii) $(A^t)^t = A$ for all m×n matrices A.

(iii) $(AB)^t = B^t A^t$ whenever the product AB is defined.

Proof

Easy exercise.

Let A be an m×n matrix whose entries are complex numbers.

The *conjugate transpose* of A is the n×m matrix with entry $\bar{a}_{ji}$ in the (i,j)-place. The conjugate transpose is denoted $\bar{A}^t$.

The conjugate transpose satisfies the same three properties as those for the transpose given in (1.1.11).

1.1.12 **The trace of a square matrix**

Let A be an n×n matrix.

We define the *trace* of A by trace $A = \sum_{i=1}^{n} a_{ii}$.

The trace of A is a single real or complex number.

1.1.13 **Proposition**

The trace has the following properties.

(i) trace (A + B) = trace A + trace B for all n×n matrices A and B.

(ii) trace $(\lambda A) = \lambda$ trace A for all n×n matrices A and all scalars λ.

(iii) trace A^t = trace A for all n×n matrices A.

(iv) trace AB = trace BA for all n×n matrices A and B.

Proof

Easy exercise to prove (i),(ii), and (iii). To prove (iv) note that the (i,i)-entry of AB is $\sum_{j=1}^{n} a_{ij}b_{ji}$ which yields that trace AB $= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji}$. Since both i and j are being summed from 1 to n this last double sum is symmetric in A and B and thus it must also give the value of trace BA.

Problems 1A

1. Let $A = \begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix}$, $B= \begin{pmatrix} 4 & -1 \\ 4 & 0 \end{pmatrix}$, $C = \begin{pmatrix} -1 & 2 \\ 2 & -1 \\ 1 & 3 \end{pmatrix}$, $D = \begin{pmatrix} 3 & 2 & 1 \\ 4 & -6 & 0 \\ 1 & -2 & -2 \end{pmatrix}$.

Calculate each of the following matrix products;

$$AB,\ CA,\ DC,\ DCAB,\ A^2,\ D^2,\ A^3B^2$$

2. Let $A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$. Prove by induction that $A^n = \begin{pmatrix} 1 & n & n(n-1) \\ 0 & 1 & 2n \\ 0 & 0 & 1 \end{pmatrix}$.

3. Let A be an m×n matrix and B an n×p matrix. Let $B_1, B_2, \ldots, B_p$ denote the columns of B. Show that $AB_1, AB_2, \ldots, AB_p$ are the columns of AB. If $A_1, A_2, \ldots, A_m$ denote the rows of A show that $A_1B, A_2B, \ldots, A_mB$ are the rows of AB.

4. Let A be an n×n matrix with entries in F. If AB = BA for all n×n matrices B with entries in F show that $A = \alpha I_n$ for some $\alpha \in F$, i.e. A is a scalar multiple of the identity matrix.

5. Let A be an n×n matrix with complex entries. If trace $\bar{A}^tA = 0$ show that A is the zero matrix.

(*Hint* - show that trace $\bar{A}^tA = \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}|^2$ where $|z|$ denotes the modulus of the complex number z.)

6. Let E_{ij} denote the n×n matrix with entry 1 in the (i,j)-place and zero elsewhere. Show that any n×n matrix $A = (a_{ij})$ is expressible in the form

$$A = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} E_{ij}.$$

Show also that $E_{ij}E_{kl} = 0$ if $j \neq k$, and $E_{ij}E_{jl} = E_{il}$.

7. Let the n×n matrix X be partitioned as follows ;

$X = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ where A is a p×p matrix, B is a p×q matrix, C is a q×p matrix, and D is a q×q matrix where p + q = n.

Let $Y = \begin{pmatrix} E & F \\ G & H \end{pmatrix}$ be an n×n matrix partitioned in a similar way. (i.e. E is a p×p matrix etc.)

Show that the product XY is partitioned as follows.

$$XY = \begin{pmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{pmatrix}$$

1.2 Systems of linear equations

A system of simultaneous linear equations

$$a_{11}x_1 + a_{12}x_2 + \ldots\ldots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \ldots\ldots + a_{2n}x_n = b_2$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \ldots\ldots + a_{mn}x_n = b_m$$

in n unknowns $x_1, x_2, \ldots, x_n$ can be rewritten as a single matrix equation $Ax = b$ where $A = (a_{ij})$ is an m×n matrix, $b = (b_i)$ is a column vector of length m, and $x = (x_i)$ is a column vector of length n.

We assume that the entries of A and b are real.

A *solution* of the system is an n-tuple of real numbers $(\alpha_1, \alpha_2, \ldots \alpha_n)$ such that $x_i = \alpha_i$ for each $i = 1,2,\ldots,n$ satisfies each of the m equations.

The *solution set* of the system is the set of all solutions of the system. It is a subset of $\mathbb{R}^n$. There are three possibilities for the solution set of the system;

(i) there is a unique solution, i.e. the solution set consists of a single point,

(ii) there are infinitely many solutions,

(iii) there are no solutions at all, i.e. the solution set is empty.

(In this case we say that the equations are *inconsistent*.)

For $m < n$ only possibilities (ii) and (iii) can occur whereas for $m \geq n$ all three possibilities can occur.

We illustrate this with a few simple examples;

1.2.1 **Example**

$$2x_1 + 3x_2 = 8$$
$$3x_1 - 3x_2 = 2$$

This system of two equations in two unknowns has the unique solution $x_1 = 2$, $x_2 = 4/3$.

Geometrically the two equations each represent a line in the plane and the solution set of the system is the point of intersection of the two lines.

1.2.2 **Example**

$$2x_1 + 3x_2 = 8$$
$$4x_1 + 6x_2 = 16$$

This system of two equations in two unknowns has infinitely many solutions. Specifically $x_1 = \alpha$, $x_2 = (8 - 2\alpha)/3$ for any $\alpha \in \mathbb{R}$ will be a solution.

Geometrically the two equations each represent the same line in the plane and the solution set of the system is the infinite set of all points on this line.

1.2.3 **Example**

$$2x_1 + 3x_2 = 8$$
$$4x_1 + 6x_2 = 3$$

This system of two equations in two unknowns has no solutions, the two equations being inconsistent.

Geometrically the two equations represent two parallel lines and so there are no points common to the two lines.

1.2.4 **Example**

$$x_1 + 2x_2 + x_3 = 3$$
$$x_1 - x_2 - x_3 = 2$$

Adding these two equations yields $2x_1 + x_2 = 5$. This gives $x_2 = 5 - 2x_1$.

Substituting into the first equation of the system then gives

$x_3 = 3 - x_1 - 2x_2 = 3 - x_1 - 2(5 - 2x_1) = 3x_1 -7$.

Thus x_1 is free to take any real number value and x_2 and x_3 are then given in terms of x_1.

The solution set is $\{ (\alpha, 5 - 2\alpha, 3\alpha - 7) \; ; \; \alpha \in \mathbb{R} \}$.

Geometrically the two equations of the system each represent a plane in $\mathbb{R}^3$ and the solution set is the line of intersection of the two planes.

1.2.5 **Remark**

In this last example m = 2, n = 3, i.e. there are more unknowns than equations. In that situation a unique solution to the system cannot be expected. There is insufficient information to be able to obtain a unique value for the unknowns so that possibility (i) cannot occur.

Geometrically two equations in three unknowns represent two planes in $\mathbb{R}^3$. These two planes can either intersect in a line as in example (1.2.4) or else be parallel and so have no points of intersection, i.e. the solution set of the corresponding system is the empty set. Similar geometric