Matrix Theory

Xingzhi Zhan

Graduate Studies in Mathematics

Volume 147

American Mathematical Society

Matrix Theory

Xingzhi Zhan

Graduate Studies
in Mathematics

Volume 147

American Mathematical Society
Providence, Rhode Island

2010 *Mathematics Subject Classification.* Primary 15-01, 15A18, 15A21, 15A60, 15A83, 15A99, 15B35, 05B20, 47A63.

For additional information and updates on this book, visit
www.ams.org/bookpages/gsm-147

Library of Congress Cataloging-in-Publication Data

Zhan, Xingzhi, 1965–
Matrix theory / Xingzhi Zhan.
pages cm — (Graduate studies in mathematics ; volume 147)
Includes bibliographical references and index.
ISBN 978-0-8218-9491-0 (alk. paper)
1. Matrices. 2. Algebras, Linear. I. Title.

QA188.Z43 2013
512.9′434—dc23

2013001353

---

10 9 8 7 6 5 4 3 2 1 18 17 16 15 14 13

Matrix Theory

Preface

The systematic study of matrices began late in the history of mathematics, but matrix theory is an active area of research now and it has applications in numerical analysis, control and systems theory, optimization, combinatorics, mathematical physics, differential equations, probability and statistics, economics, information theory, and engineering.

One attractive feature of matrix theory is that many matrix problems can be solved naturally by using tools or ideas from other branches of mathematics such as analysis, algebra, graph theory, geometry and topology. The reverse situation also occurs, as shown in the last chapter.

This book is intended for use as a text for graduate or advanced undergraduate level courses, or as a reference for research workers. It is based on lecture notes for graduate courses I have taught five times at East China Normal University and once at Peking University. My aim is to provide a concise treatment of matrix theory. I hope the book contains the basic knowledge and conveys the flavor of the subject.

When I chose material for this book, I had the following criteria in mind: 1) important; 2) elegant; 3) ingenious; 4) interesting. Of course, a very small percentage of mathematics meets all of these criteria, but I hope the results and proofs here meet at least one of them. As a reader I feel that for clarity, the logical steps of a mathematical proof cannot be omitted, though routine calculations may be or should be. Whenever possible, I try to have a conceptual understanding of a result. I always emphasize methods and ideas.

Most of the exercises are taken from research papers, and they have some depth. Thus if the reader has difficulty in solving the problems in these exercises, she or he should not feel frustrated.

Parts of this book appeared in a book in Chinese with the same title published by the Higher Education Press in 2008.

Thanks go to Professors Pei Yuan Wu and Wei Wu for discussions on the topic of numerical range and to Dr. Zejun Huang for discussions on Theorem 1.2 and Lemma 9.13. I am grateful to Professors Tsuyoshi Ando, Rajendra Bhatia, Richard Brualdi, Roger Horn, Erxiong Jiang, Chi-Kwong Li, Zhi-Guo Liu, Jianyu Pan, Jia-Yu Shao, Sheng-Li Tan, and Guang Yuan Zhang for their encouragement, friendship and help over the years. I wish to express my gratitude to my family for their kindness. This work was supported by the National Science Foundation of China under grant 10971070.

Shanghai, December 2012 Xingzhi Zhan

Contents

Chapter 1

Preliminaries

Most of the concepts and results in this chapter will be used in the sequel. We also set up some notation.

We mainly consider complex matrices which include real matrices, of course. Occasionally we deal with matrices over a generic field. A *square matrix* is a matrix that has the same number of rows and columns, while a *rectangular matrix* is a matrix the numbers of whose rows and columns may be unequal. An $m \times n$ matrix is a matrix with m rows and n columns. An $n \times n$ matrix is said to be of *order* n. An $m \times 1$ matrix is called a *column vector*, and a $1 \times n$ matrix is called a *row vector*. Thus vectors are special matrices.

A matrix over a set Ω means that its entries are elements of Ω. Usually the set Ω is a field or a ring. We denote by $M_{m,n}(\Omega)$ the set of the $m \times n$ matrices over Ω. Here the letter M suggests matrix. $M_{n,n}(\Omega)$ will be abbreviated as $M_n(\Omega)$. When $\Omega = \mathbb{C}$, the field of complex numbers, $M_{m,n}(\mathbb{C})$ and $M_n(\mathbb{C})$ are simply written as $M_{m,n}$ and M_n, respectively. Ω^n denotes the set of n-tuples with components from Ω. Unless otherwise stated, the elements of Ω^n are written in the form of column vectors so that they can be multiplied by matrices on the left.

If A is a matrix, $A(i,j)$ denotes its entry in the i-th row and j-th column. We say that this entry is in the position (i,j). The notation $A = (a_{ij})_{m \times n}$ means that A is an $m \times n$ matrix with $A(i,j) = a_{ij}$. A^T denotes the transpose of a matrix A. If $A \in M_{m,n}$, $\bar{A}$ denotes the matrix obtained from A by taking the complex conjugate entrywise, and A^* denotes the conjugate transpose of A, i.e., $A^* = (\bar{A})^T$. Thus, if x is a column vector, then x^T and x^* are row vectors. For simplicity, we use 0 to denote the zero matrix, and we use

I to denote the identity matrix, i.e., the diagonal matrix with all diagonal entries being 1. Their sizes will be clear from the context.

Denote by $\text{diag}(d_1, \ldots, d_n)$ the diagonal matrix with diagonal entries $d_1, \ldots, d_n$. If A_i, $i = 1, \ldots, k$ are square matrices, sometimes we use the notation $A_1 \oplus A_2 \oplus \cdots \oplus A_k$ to denote the block diagonal matrix

$$\text{diag}(A_1, A_2, \ldots, A_k) = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_k \end{bmatrix}.$$

We will use $G \triangleq \cdots$ to mean that we define G to be something. This notation can streamline the presentation. ϕ will denote the empty set, unless otherwise stated.

1.1. Classes of Special Matrices

Let $A \in M_n$. If $A^*A = AA^*$, then A is called *normal*. If $A^* = A$, then A is called *Hermitian*. If $A^* = -A$, then A is called *skew-Hermitian*. If $A^*A = I$, then A is called *unitary*. Thus, unitary matrices are those matrices A satisfying $A^{-1} = A^*$. Obviously, Hermitian matrices, skew-Hermitian matrices, and unitary matrices are normal matrices; real Hermitian matrices are just real symmetric matrices, and real unitary matrices are just real orthogonal matrices. The set of all the eigenvalues of a square complex matrix A is called the *spectrum* of A, and is denoted by $\sigma(A)$. Note that $\sigma(A)$ is a multi-set if A has repeated eigenvalues. The *spectral radius* of A is defined and denoted by $\rho(A) = \max\{|\lambda| : \lambda \in \sigma(A)\}$.

Theorem 1.1 (Spectral Decomposition). *Every normal matrix is unitarily similar to a diagonal matrix; i.e., if $A \in M_n$ is normal, then there exists a unitary matrix $U \in M_n$ such that*

$$A = U\text{diag}(\lambda_1, \ldots, \lambda_n)U^*. \tag{1.1}$$

Obviously $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A, and they can appear on the diagonal in any prescribed order.

We will prove this theorem in Section 1.7 using Schur's unitary triangularization theorem.

Denote by $\langle \cdot, \cdot \rangle$ the standard Euclidean inner product on $\mathbb{C}^n$. If $x = (x_1, \ldots, x_n)^T$, $y = (y_1, \ldots, y_n)^T \in \mathbb{C}^n$, then $\langle x, y \rangle = \sum_{j=1}^{n} x_j \overline{y}_j = y^*x$. The vector space $\mathbb{C}^n$ with this inner product is a Hilbert space. A matrix $A \in M_n$ may be regarded as a linear operator on $\mathbb{C}^n : x \mapsto Ax$. Since $\langle Ax, y \rangle =$

$\langle x, A^*y\rangle$ for all $x, y \in \mathbb{C}^n$, the conjugate transpose A^* is exactly the *adjoint* of A in the operator theory setting.

$A \in M_n$ is said to be *positive semidefinite* if

$$\langle Ax, x\rangle \geq 0 \quad \text{for all} \quad x \in \mathbb{C}^n. \tag{1.2}$$

$A \in M_n$ is said to be *positive definite* if

$$\langle Ax, x\rangle > 0 \quad \text{for all} \quad 0 \neq x \in \mathbb{C}^n. \tag{1.3}$$

Positive definite matrices are exactly invertible positive semidefinite matrices. Invertible matrices are also called *nonsingular*, while square matrices that have no inverse are called *singular*.

As usual, denote by $\mathbb{R}$ the field of real numbers. For $A \in M_n$ and $x, y \in \mathbb{C}^n$ we have the following *polarization identities:*

$$4\langle Ax, y\rangle = \sum_{k=0}^{3} i^k \langle A(x+i^k y), x+i^k y\rangle,$$
$$4\langle x, Ay\rangle = \sum_{k=0}^{3} i^k \langle x+i^k y, A(x+i^k y)\rangle,$$

where $i = \sqrt{-1}$. It follows from these two identities that for a given $A \in M_n$, if $\langle Ax, x\rangle \in \mathbb{R}$ for any $x \in \mathbb{C}^n$, then A is Hermitian. In particular, the defining condition (1.2) implies that a positive semidefinite matrix is necessarily Hermitian. In fact, positive semidefinite matrices are those Hermitian matrices which have nonnegative eigenvalues, and positive definite matrices are those Hermitian matrices which have positive eigenvalues. If $A \in M_n$ is positive semidefinite, then for any $B \in M_{n,k}$, B^*AB is positive semidefinite; if $A \in M_n$ is positive definite, then for any nonsingular $B \in M_n$, B^*AB is positive definite.

Let $A \in M_n(\mathbb{R})$ be real and symmetric. For $x \in \mathbb{C}^n$, let $x = y + iz$ where $y, z \in \mathbb{R}^n$. Then $x^*Ax = y^TAy + z^TAz$. Thus a real symmetric matrix A is positive semidefinite if and only if $x^TAx \geq 0$ for all $x \in \mathbb{R}^n$, and it is positive definite if and only if $x^TAx > 0$ for all $0 \neq x \in \mathbb{R}^n$.

A matrix is said to be *diagonalizable* if it is similar to a diagonal matrix. From the Jordan canonical form it is clear that if $A \in M_n$ has n distinct eigenvalues, then A is diagonalizable.

$A = (a_{ij}) \in M_n$ is called *upper triangular* if $a_{ij} = 0$ for all $i > j$, i.e., the entries below the diagonal are zero. If $a_{ij} = 0$ for all $i \geq j$ then A is called *strictly upper triangular.*

$A = (a_{ij}) \in M_n$ is called *lower triangular* if $a_{ij} = 0$ for all $i < j$, i.e., the entries above the diagonal are zero. If $a_{ij} = 0$ for all $i \leq j$ then A is called *strictly lower triangular.*

It is easy to verify that the product of two upper (lower) triangular matrices is upper (lower) triangular and the inverse of an upper (lower) triangular matrix is upper (lower) triangular.

$A = (a_{ij}) \in M_n$ is called a *Hessenberg matrix* if $a_{ij} = 0$ for all $i > j + 1$.

We say that a matrix $A = (a_{ij})$ has *upper bandwidth* p if $a_{ij} = 0$ for all i, j with $j - i > p$; A has *lower bandwidth* q if $a_{ij} = 0$ for all i, j with $i - j > q$. For example, lower triangular matrices have upper bandwidth 0, and Hessenberg matrices have lower bandwidth 1. A matrix $A \in M_n$ is called a *band matrix* if A has upper bandwidth $p \leq n - 2$ or has lower bandwidth $q \leq n - 2$.

A matrix is called a *sparse matrix* if it has many zero entries. This is not a precise notion.

$A = (a_{ij}) \in M_{m,n}$ is called a *0-1 matrix* if every entry $a_{ij} \in \{0, 1\}$. A square 0-1 matrix that has exactly one 1 in each row and in each column is called a *permutation matrix.*

$A = (a_{ij}) \in M_n$ is called a *Toeplitz matrix* if there are numbers

$$a_{-n+1}, \ldots, a_{-1}, a_0, a_1, \ldots, a_{n-1}$$

such that $a_{ij} = a_{j-i}$. Hence a Toeplitz matrix is a matrix of the form

$$\begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{n-1} \\ a_{-1} & a_0 & a_1 & \cdots & a_{n-2} \\ a_{-2} & a_{-1} & a_0 & \cdots & a_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{-n+1} & a_{-n+2} & a_{-n+3} & \cdots & a_0 \end{bmatrix}.$$

$A = (a_{ij}) \in M_n$ is called a *Hankel matrix* if there are numbers $a_1, \ldots, a_{2n-1}$ such that $a_{ij} = a_{i+j-1}$. Hence a Hankel matrix is a matrix of the form

$$\begin{bmatrix} a_1 & a_2 & a_3 & \cdots & a_n \\ a_2 & a_3 & a_4 & \cdots & a_{n+1} \\ a_3 & a_4 & a_5 & \cdots & a_{n+2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_n & a_{n+1} & a_{n+2} & \cdots & a_{2n-1} \end{bmatrix}.$$

A matrix of the form

$$A = \begin{bmatrix} a_1 & a_2 & a_3 & \cdots & a_n \\ a_n & a_1 & a_2 & \cdots & a_{n-1} \\ a_{n-1} & a_n & a_1 & \cdots & a_{n-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_2 & a_3 & a_4 & \cdots & a_1 \end{bmatrix} \tag{1.4}$$

is called a *circulant matrix.* Such an A is determined by the first row, and each row is just the previous row cycled forward one step. Denote the matrix in (1.4) by $\mathrm{Circ}(a_1, a_2, a_3, \ldots, a_n)$. $P = \mathrm{Circ}(0, 1, 0, \ldots, 0)$ is called the *basic circulant matrix.* Note that P is a permutation matrix. We have

$$A = \sum_{k=0}^{n-1} a_{k+1} P^k. \tag{1.5}$$

The characteristic polynomial of P is $\lambda^n - 1$, so its eigenvalues are $z^j, j = 0, 1, \ldots, n-1$, where $z = e^{\frac{2\pi}{n}i}, i = \sqrt{-1}$. Let $x_j = \frac{1}{\sqrt{n}}(1, z^j, z^{2j}, \ldots, z^{(n-1)j})^T$. Then $x_0, x_1, \ldots, x_{n-1}$ are orthonormal eigenvectors of P. Let $U = (x_0, x_1, \ldots, x_{n-1})$. Then U is a unitary matrix and

$$P = U \mathrm{diag}(1, z, z^2, \ldots, z^{n-1}) U^*. \tag{1.6}$$

Let $f(t) = \sum_{k=0}^{n-1} a_{k+1} t^k$. (1.5) and (1.6) yield

$$A = U \mathrm{diag}(f(1), f(z), f(z^2), \ldots, f(z^{n-1})) U^*. \tag{1.7}$$

(1.7) shows that all the circulant matrices in M_n can be unitarily diagonalized by one fixed unitary matrix.

A matrix of the form

$$V = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ a_1 & a_2 & a_3 & \cdots & a_n \\ a_1^2 & a_2^2 & a_3^2 & \cdots & a_n^2 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ a_1^{n-1} & a_2^{n-1} & a_3^{n-1} & \cdots & a_n^{n-1} \end{bmatrix}$$

is called a *Vandermonde matrix.* Since the determinant

$$\det V = \prod_{i>j} (a_i - a_j),$$

V is nonsingular if and only if $a_1, a_2, \ldots, a_n$ are distinct. The Vandermonde matrix has several variants. For example, a matrix of the form

$$W = \begin{bmatrix} a_1^{n-1} & a_1^{n-2} & \cdots & a_1 & 1 \\ a_2^{n-1} & a_2^{n-2} & \cdots & a_2 & 1 \\ a_3^{n-1} & a_3^{n-2} & \cdots & a_3 & 1 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_n^{n-1} & a_n^{n-2} & \cdots & a_n & 1 \end{bmatrix}$$

is also called a Vandermonde matrix whose determinant is

$$\det W = \prod_{i<j} (a_i - a_j).$$

1.2. The Characteristic Polynomial

By definition, the *characteristic polynomial* of a square matrix A is $f(t) = \det(tI - A)$.

Theorem 1.2. *Let $E_k(A)$ be the sum of all the $k \times k$ principal minors of a matrix A of order n over a field. Then the characteristic polynomial of A is*

$$f(t) = t^n - E_1(A)t^{n-1} + E_2(A)t^{n-2} - \cdots + (-1)^n E_n(A). \tag{1.8}$$

Proof. For a matrix G of order n and integer indices $1 \le i_1 < i_2 < \cdots < i_k \le n$, we denote by $G(i_1, i_2, \ldots, i_k)$ the principal submatrix of G obtained from G by deleting rows $i_1, i_2, \ldots, i_k$ and deleting columns $i_1, i_2, \ldots, i_k$. Introduce indeterminates $t_1, \ldots, t_n$ and consider the matrix

$$B = \operatorname{diag}(t_1, t_2, \ldots, t_n) - A.$$

We have the following expansion of the determinant:

$$\begin{aligned} \det B = \prod_{j=1}^{n} t_j - \sum_{1 \le i_1 < i_2 < \cdots < i_{n-1} \le n} \det A(i_1, i_2, \ldots, i_{n-1}) \prod_{j=1}^{n-1} t_{i_j} \\ + \sum_{1 \le i_1 < i_2 < \cdots < i_{n-2} \le n} \det A(i_1, i_2, \ldots, i_{n-2}) \prod_{j=1}^{n-2} t_{i_j} - \cdots + (-1)^n \det A. \end{aligned} \tag{1.9}$$

To see this, first note that the constant term in the expansion of $\det B$ is $(-1)^n \det A = (-1)^n E_n(A)$. Then for any but fixed indices $1 \le i_1 < i_2 < \cdots < i_k \le n$, using the Laplace expansion according to the rows $i_1, i_2, \ldots, i_k$ we see that the coefficient of $t_{i_1} t_{i_2} \cdots t_{i_k}$ is $(-1)^{n-k} \det A(i_1, i_2, \ldots, i_k)$, since the constant term in the expansion of $\det B(i_1, i_2, \ldots, i_k)$ is

$$(-1)^{n-k} \det A(i_1, i_2, \ldots, i_k).$$

This proves (1.9).

Now in B, setting $t_1 = t_2 = \cdots = t_n = t$ we get (1.8), since

$$E_{n-k}(A) = \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} \det A(i_1, i_2, \ldots, i_k).$$

□

The following theorem gives a basic relation between the coefficients of a polynomial and the moments of its roots. Note that a polynomial of degree n over a field F has n roots (including multiplicities) in the algebraic closure of F.

Theorem 1.3. *Let $f(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$ be a monic polynomial over a field with roots $\lambda_1, \ldots, \lambda_n$, including multiplicities. Denote the k-th*