Perturbation Bounds for Matrix Eigenvalues

Rajendra Bhatia

C • L • A • S • S • I • C • S
In Applied Mathematics
53

siam.

Perturbation Bounds for Matrix Eigenvalues

Books in the Classics in Applied Mathematics series are monographs and textbooks declared out of print by their original publishers, though they are of continued importance and interest to the mathematical community. SIAM publishes this series to ensure that the information presented in these texts is not lost to today's students and researchers.

Classics in Applied Mathematics

C. C. Lin and L. A. Segel, *Mathematics Applied to Deterministic Problems in the Natural Sciences*
Johan G. F. Belinfante and Bernard Kolman, *A Survey of Lie Groups and Lie Algebras with Applications and Computational Methods*
James M. Ortega, *Numerical Analysis: A Second Course*
Anthony V. Fiacco and Garth P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*
F. H. Clarke, *Optimization and Nonsmooth Analysis*
George F. Carrier and Carl E. Pearson, *Ordinary Differential Equations*
Leo Breiman, *Probability*
R. Bellman and G. M. Wing, *An Introduction to Invariant Imbedding*
Abraham Berman and Robert J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*
Olvi L. Mangasarian, *Nonlinear Programming*
*Carl Friedrich Gauss, *Theory of the Combination of Observations Least Subject to Errors: Part One, Part Two, Supplement.* Translated by G. W. Stewart
Richard Bellman, *Introduction to Matrix Analysis*
U. M. Ascher, R. M. M. Mattheij, and R. D. Russell, *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations*
K. E. Brenan, S. L. Campbell, and L. R. Petzold, *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*
Charles L. Lawson and Richard J. Hanson, *Solving Least Squares Problems*
J. E. Dennis, Jr. and Robert B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*
Richard E. Barlow and Frank Proschan, *Mathematical Theory of Reliability*
Cornelius Lanczos, *Linear Differential Operators*
Richard Bellman, *Introduction to Matrix Analysis, Second Edition*

*First time in print.

Classics in Applied Mathematics (continued)

Beresford N. Parlett, *The Symmetric Eigenvalue Problem*

Richard Haberman, *Mathematical Models: Mechanical Vibrations, Population Dynamics, and Traffic Flow*

Peter W. M. John, *Statistical Design and Analysis of Experiments*

Tamer Başar and Geert Jan Olsder, *Dynamic Noncooperative Game Theory, Second Edition*

Emanuel Parzen, *Stochastic Processes*

Petar Kokotović, Hassan K. Khalil, and John O'Reilly, *Singular Perturbation Methods in Control: Analysis and Design*

Jean Dickinson Gibbons, Ingram Olkin, and Milton Sobel, *Selecting and Ordering Populations: A New Statistical Methodology*

James A. Murdock, *Perturbations: Theory and Methods*

Ivar Ekeland and Roger Témam, *Convex Analysis and Variational Problems*

Ivar Stakgold, *Boundary Value Problems of Mathematical Physics, Volumes I and II*

J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*

David Kinderlehrer and Guido Stampacchia, *An Introduction to Variational Inequalities and Their Applications*

F. Natterer, *The Mathematics of Computerized Tomography*

Avinash C. Kak and Malcolm Slaney, *Principles of Computerized Tomographic Imaging*

R. Wong, *Asymptotic Approximations of Integrals*

O. Axelsson and V. A. Barker, *Finite Element Solution of Boundary Value Problems: Theory and Computation*

David R. Brillinger, *Time Series: Data Analysis and Theory*

Joel N. Franklin, *Methods of Mathematical Economics: Linear and Nonlinear Programming, Fixed-Point Theorems*

Philip Hartman, *Ordinary Differential Equations, Second Edition*

Michael D. Intriligator, *Mathematical Optimization and Economic Theory*

Philippe G. Ciarlet, *The Finite Element Method for Elliptic Problems*

Jane K. Cullum and Ralph A. Willoughby, *Lanczos Algorithms for Large Symmetric Eigenvalue Computations, Vol. I: Theory*

M. Vidyasagar, *Nonlinear Systems Analysis, Second Edition*

Robert Mattheij and Jaap Molenaar, *Ordinary Differential Equations in Theory and Practice*

Shanti S. Gupta and S. Panchapakesan, *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations*

Eugene L. Allgower and Kurt Georg, *Introduction to Numerical Continuation Methods*

Leah Edelstein-Keshet, *Mathematical Models in Biology*

Heinz-Otto Kreiss and Jens Lorenz, *Initial-Boundary Value Problems and the Navier-Stokes Equations*

J. L. Hodges, Jr. and E. L. Lehmann, *Basic Concepts of Probability and Statistics, Second Edition*

George F. Carrier, Max Krook, and Carl E. Pearson, *Functions of a Complex Variable: Theory and Technique*

Friedrich Pukelsheim, *Optimal Design of Experiments*

Israel Gohberg, Peter Lancaster, and Leiba Rodman, *Invariant Subspaces of Matrices with Applications*

Lee A. Segel with G. H. Handelman, *Mathematics Applied to Continuum Mechanics*

Rajendra Bhatia, *Perturbation Bounds for Matrix Eigenvalues*

Perturbation Bounds for Matrix Eigenvalues

Rajendra Bhatia

siam.
Society for Industrial and Applied Mathematics
Philadelphia

This SIAM edition is an unabridged republication of the work first published by Longman Scientific & Technical, 1987.

10 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging-in-Publication Data:

Bhatia, Rajendra, 1952-
Perturbation bounds for matrix eigenvalues / Rajendra Bhatia
p. cm. — (Classics in applied mathematics ; 53)
Originally published: Burnt Mill, Harlow, Essex, England : Longman Scientific & Technical ; New York : Wiley, 1987.
ISBN 978-0-898716-31-3 (alk. paper)
1. Eigenvalues. 2. Perturbation (Mathematics) 3. Matrices. I. Title.

QA193.B48 2007
512.9'436—dc22

2007061736

Contents

Supplements 1986-2006

Preface to the Classics Edition

The first version of this book was written in 1985 and published in 1987. Its principal goal was the exposition of bounds for the distance between the eigenvalues of two matrices A and B in terms of expressions involving $\|A - B\|$. The prototype of such bounds is H. Weyl's inequality from 1912.

This subject flourished in the 1950s, with important contributions by A. M. Ostrowski, V. B. Lidskii, A. J. Hoffman, H. W. Wielandt, and others. Three noteworthy papers were added to these in the early 1960s. One of them, by L. Mirsky (1960), was an illuminating survey article and it formulated some problems concerning normal matrices. The other two by F. L. Bauer and C. T. Fike (1960) and by P. Henrici (1962) focused on nonnormal matrices. Then for several years not much was added to *this* aspect of the subject, even as the venerable books by T. Kato and J. H. Wilkinson appeared in 1965–66.

A revival, in which the predominant role was played by my coworkers, occurred around 1980. It was a happy coincidence that the appearance of Marshall and Olkin's book just at this time kindled widespread interest in majorization. As a result, lots of matrix inequalities, among them several perturbation bounds, were discovered in the 1980s.

This book, written as these developments were taking place, attempted to present a unified picture of the old and the new results. It seems to have had more success than I could have imagined at that time, capturing the attention of several prominent practitioners of the subject, stimulating further work, and receiving favorable reviews. I am grateful for this response.

Appearing, as it did, in a series of "Research Notes" that are considered to be of transient value, the book went out of print three years after its publication. I am both pleased and honored that, twenty years later, the editors at SIAM have considered it appropriate to republish this book in the series Classics in Applied Mathematics.

In these intervening years, a lot of work has taken place in the subject. To maintain the value of the book as a research resource I have added an appendix entitled "Supplements 1986–2006." This consists of seven parts, five of which supplement Chapters 2 to 6 of the original book, and have been given the same chapter titles for the convenience of the reader. The supplements are divided into sections numbered 26 to 43 in continuity with the 25 sections in the first edition. Again I have chosen to include only those results that have the same form as the prototype Weyl's inequality. The book can be used for a short "topics" course in linear algebra or functional analysis at the graduate level. Keeping that use in mind, the supplement to Chapter 2 even contains a bit of classical material that is used in later sections.

The introduction in the first edition contained a summary of the important inequalities presented in the book. This could have been the prompt for a kind reviewer to say that "the monograph should also prove a boon for those who merely want to use it quickly and run." The quick and the patient reader both may appreciate being informed about prominent changes in the subject since that introduction was written. I list them in brief in the next paragraphs.

For several years the most prominent conjecture on perturbation inequalities, which attracted the attention of several mathematicians, was that the inequality (3) in the introduction,

$$d(\operatorname{Eig} A,\ \operatorname{Eig} B)\ \leq\ \|A-B\|,$$

would be true for all normal matrices A and B. In 1992, J. Holbrook published a counterexample to this with 3×3 matrices. It is now known that the inequality (4)

$$d(\operatorname{Eig} A,\ \operatorname{Eig} B)\ \leq\ c\,\|A-B\|$$

is true for all $n \times n$ normal matrices A and B with $c < 2.904$ and that the best constant c here is bigger than 1.018.

The inequality (9) of the introduction has been significantly improved. The factor n occurring there can be replaced by a small number independent of n. For any two $n \times n$ matrices A and B we have

$$d(\operatorname{Eig} A,\ \operatorname{Eig} B)\ \leq\ 4\,(2M)^{1-1/n}\ \|A-B\|^{1/n}.$$

This is a consequence of work done by D. Phillips and by R. Bhatia, L. Elsner, and G. Krause in 1990.

If A is normal and B arbitrary, the inequality (8) can be improved to

$$d(\operatorname{Eig} A,\ \operatorname{Eig} B) \leq n\, \|A - B\|.$$

This was shown by J.-G. Sun in 1996. In Section 40 the reader will find inequalities for other norms in this situation.

The theorems in Chapters 3 and 4 have been significantly generalized in various directions. The new Chapter 10 deals with pairs of diagonalizable matrices. These are matrices A and B such that $A = SD_1S^{-1}$ and $B = TD_2T^{-1}$, where D_1 and D_2 are diagonal, and S and T are invertible matrices. The number $c(S) = \|S\|\,\|S^{-1}\|$ is called the condition number of S. When A is normal, S can be chosen to be unitary, and then $c(S) = 1$. Suppose the diagonal matrices D_1 and D_2 are real. Then we have

$$|||\operatorname{Eig}_{\downarrow}(A) - \operatorname{Eig}_{\downarrow}(B)||| \leq \sqrt{c(S)c(T)}\ |||A - B|||.$$

This is a pleasant generalization of the inequality (12) in the introduction. Exactly the same type of extension of the inequality (16) has been established for arbitrary diagonalizable matrices, and of inequalities (3) and (14) for diagonalizable matrices whose spectra are on the unit circle.

In another direction these theorems on Hermitian, unitary, and normal operators have been extended to the case when A and B are operators on an infinite-dimensional Hilbert space with the restriction that $A - B$ is compact. These results are summarized in Section 41.

The somewhat daunting proof of Lidskii's Theorem in Chapter 3 can be substantially simplified by an argument of C.-K. Li and R. Mathias. This is presented in Section 30. One of the most spectacular developments in recent years has been the solution of Horn's problem. This is related to our discussion but is not our main theme. Another important advance that has taken place is the proof of Lax's conjecture on hyperbolic polynomials. A direct consequence of this for our problem is that when A and B are matrices all whose real linear combinations have only real eigenvalues, then not only the inequality (1) but also the inequality (12) of the introduction holds good.

I am thankful to Jim Demmel, Ludwig Elsner, Leon Gurvits, John Holbrook, Roger Horn, Adrian Lewis, Chi-Kwong Li, Ren-Cang Li, Roy

Mathias, and Xingzhi Zhan, all of whom responded to my request for information that I needed for preparing the supplements.

The proposal to reprint this book originated from series editor Nick Higham. I thank him for this interest in giving the book a new life, and I thank the editors and the staff at SIAM for their help and support.

Preface

These notes are based on lectures that I gave at Hokkaido University in the fall of 1985. Substantial parts of this material were presented earlier in seminars at the University of Bombay, the University of Toronto and the Indian Statistical Institute. It is a pleasure to record my thanks to Professors T. Ando, Ch. Davis and M.G. Nadkarni who organized these seminars.

Professors F. Hiai and K.R. Parthasarathy were kind enough to go through my handwritten notes. Their observations led to the elimination of several errors. The ones that have survived are, of course, solely my responsibility

I thank the editorial staff of Longman Scientific and Technical for their typing of this monograph in its present form and for its production.

The library resources available to me while preparing these notes were somewhat limited. It is, therefore, likely that the work of some authors has been misrepresented or overlooked. I apologise for all such omissions.

Rajendra Bhatia

Introduction

Introduction

Ever since the publication of the classic *Methods of Mathematical Physics* by Courant and Hilbert, eigenvalues have occupied a central position in applied mathematics and engineering. Vibrations occur everywhere in nature; each vibration has a certain frequency; these frequencies are the eigenvalues of some differential operator describing the physical system.

Finding the eigenvalues of an operator is not always an easy task. Sometimes it is easier to calculate the eigenvalues of a nearby operator and then use this knowledge to locate approximately the eigenvalues of the original operator. In some problems, the underlying physical system may be subjected to changes (perturbations) and we may want to determine the consequent change in eigenvalues. On other occasions, we may know an operator only approximately due to errors of observation, or we may have to feed an approximation of it to a computing device. In each case we would like to know how much this error or approximation would affect the eigenvalues of the operator.

All these considerations give rise to one mathematical problem: if we know how close two operators are, can we say how close their eigenvalues must be? It is to this problem - or rather to one of the several facets of this problem - that the following pages are devoted.

First of all we restrict ourselves here to the study of finite-dimensional operators (matrices). The finite-dimensional theory from the mathematical physicist's point of view has been dealt with in the monographs of Baumgärtel [1] and Kato [1]. The central question here is: if A(t) is a family of matrices varying smoothly with a parameter t then do the eigenvalues, eigenvectors, eigenprojections and eigennilpotents of A(t) also vary smoothly with t? If so, what are the power series expansions, their radii of convergence and the error estimates when the power series are truncated? From the numerical analyst's point of view, the theory has been expounded in the book of Wilkinson [1] and its successor, the recent work of Parlett [1]. Here the emphasis is on actual methods of computation, the rates of convergence of various algorithms and how to accelerate them, how to use one part of a computation to simplify the remaining. One of the many topics dealt with in all these books is finding good error bounds or perturbation inequalities for eigenvalues. That is the only topic we will study here, though in greater detail.

We further restrict ourselves to the study of only one kind of perturbation inequalities. Our object here would be global, *a priori* bounds for the distance between the eigenvalue n-tuples of two n×n matrices. Thus, we would not require that the two matrices should be close to each other by a preassigned amount; we would not require or use any knowledge about one part of the spectrum (or any of the associated objects like the Jordan form) of one of the operators; and we would deal with all the eigenvalues at the same time. We must emphasize that in practical situations some of this knowledge is readily available, either to begin with or at some intermediate stage of a calculation. So it may be unwise to discard it. Nevertheless, we will assume that we are given only the size of the difference, $\|A-B\|$, for two matrices A and B and perhaps that A,B are from some special class of matrices.

We now give a brief summary of the major inequalities which are proved (occasionally just stated) in this monograph. This could serve as a preview for some readers, while others just looking for an inequality to use would be saved the trouble of digging it out from the text.

The prototype of (almost) all of our inequalities is the following result of H. Weyl (1912). Let A,B be Hermitian matrices with eigenvalues $\alpha_1 \geq \dots \geq \alpha_n$ and $\beta_1 \geq \dots \geq \beta_n$, respectively. Then

$$\max_j |\alpha_j - \beta_j| \leq \|A-B\| . \qquad (1)$$

Here, $\|A\|$ denotes the operator bound norm of A.

P. Lax (1958) showed that the same result is true when A,B are any two matrices all whose real linear combinations have only real eigenvalues. Such matrices arise in the study of hyperbolic partial differential equations. (In fact, Lax's proof of this result relied on methods of partial differential equations).

If A,B are arbitrary n×n matrices with eigenvalues $\alpha_1,\dots,\alpha_n$ and $\beta_1,\dots,\beta_n$, respectively, define the optimal matching distance between their eigenvalues as

$$d(\text{Eig } A, \text{Eig } B) = \min_{\sigma} \max_{j} |\alpha_j - \beta_{\sigma(j)}|, \tag{2}$$

where the minimum is taken over all permutations on n symbols.

R. Bhatia and C. Davis (1984) showed that if A,B are unitary matrices then

$$d(\text{Eig } A, \text{Eig } B) \leq \|A-B\|. \tag{3}$$

Note that, if A,B are Hermitian, then the left hand side of (1) is equal to d(Eig A, Eig B). So this result is analogous to that of Weyl.

It has long been conjectured that the inequality (3) remains true when A,B are normal matrices. Only some special cases of this conjecture have been proved so far. It is true when A,B and A-B are all normal (R.Bhatia (1982)), when A,B are constant multiples of unitaries (R.Bhatia and J.A.R. Holbrook (1985)) and when A is Hermitian and B skew-Hermitian (V.S. Sunder (1982)).

More interesting, perhaps, is the fact that a weaker form of the conjecture has been established. R. Bhatia, C. Davis and A. McIntosh (1983) showed that if A,B are normal matrices then

$$d(\text{Eig } A, \text{Eig } B) \leq c\|A-B\| \tag{4}$$

where c is a universal constant independent of the dimension. The value of this constant is not known, but Koosis (unpublished) has shown that $c < \pi$.

W. Kahan (1975) showed that if A is Hermitian and B arbitrary then

$$d(\text{Eig } A, \text{Eig } B) \leq (\gamma_n + 2)\ \|A-B\| \tag{5}$$

where γ_n is a constant depending on the size n of the matrices. Further he showed that the optimal constant for this inequality is bounded as

$$\frac{2}{\pi} \ln n - O(1) \leq \gamma_n \leq \log_2 n + 0.038. \tag{6}$$

The value of this constant was found by A. Pokrzywa (1981) who showed

$$\gamma_n = \frac{2}{n} \sum_{j=1}^{[n/2]} \cot \frac{2j-1}{2n} \pi. \tag{7}$$

One can see that if A is normal and B arbitrary then

$$d(\text{Eig } A, \text{ Eig } B) \le (2n-1) \ \|A-B\|. \tag{8}$$

If, in addition, B is Hermitian then the factor (2n-1) can be replaced by $\sqrt{2}$ in the above inequality.

When A,B are arbitrary nxn matrices the situation is not so simple. Results of this type in the general case were obtained by A. Ostrowski (1957), P. Henrici (1962), R. Bhatia and K.K. Mukherjea (1979), R. Bhatia and S. Friedland (1981) and L. Elsner (1982) and (1985). This latest result of Elsner says that for A,B arbitrary nxn matrices

$$d(\text{Eig } A, \text{ Eig } B) \le n \ (2M)^{1-1/n} \ \|A-B\|^{1/n}, \tag{9}$$

where, $M = \max(\|A\|, \ \|B\|)$.

Let us now turn to another kind of generalization of Weyl's inequality (1). All through the above discussion we used the operator bound norm to measure the size of an operator. For some problems this norm is not very suitable. For one thing, it is not always easy to compute it; for another, it is not always possible to assert that an operator is uniformly small on the entire space (which is what the assertion $\|A\| < \varepsilon$ would mean). So, it is often advantageous to obtain estimates in other norms as well. Of particular geometrical interest are the "unitarily-invariant norms", like the Frobenius norm

$$\|A\|_F = (\text{tr } A^*A)^{1/2} = (\sum_{i,j} |a_{ij}|^2)^{1/2} \tag{10}$$

or the trace norm

$$\|A\|_{tr} = \text{tr}(A^*A)^{1/2}. \tag{11}$$

Unitary-invariance here means the property $\|A\| = \|UAV\|$, for any two unitary matrices U and V. We will denote by $|||\cdot|||$ any of the family of unitarily-invariant norms, and the statement $|||A||| \leq |||B|||$ would mean that this inequality is true for all these norms simultaneously.

Let A be a Hermitian matrix with eigenvalues $\alpha_1 \geq \dots \geq \alpha_n$ and denote by $\mathrm{Eig}_{\downarrow}(A)$ the diagonal matrix with $\alpha_1,\dots,\alpha_n$ as its diagonal entries. Then the inequality (1) can be restated as

$$\|\mathrm{Eig}_{\downarrow}(A) - \mathrm{Eig}_{\downarrow}(B)\| \leq \|A-B\|.$$

From a theorem of V. Lidskii (1950) together with von Neumann's characterization of unitarily-invariant norms and some inequality results of Hardy, Littlewood and Polya, one obtains a big generalization of this. We have for A,B Hermitian

$$|||\mathrm{Eig}_{\downarrow}(A) - \mathrm{Eig}_{\downarrow}(B)||| \leq |||A-B|||. \tag{12}$$

For arbitrary matrices A,B we can define a distance between their eigenvalues in any norm as follows. Denote by Eig A the diagonal matrix with the eigenvalues of A placed on the diagonal in any order. Then define

$$|||(\mathrm{Eig}\ A,\ \mathrm{Eig}\ B)||| = \min_{W} |||\mathrm{Eig}\ A - W(\mathrm{Eig}\ B)W^{-1}|||, \tag{13}$$

where W runs over all permutation matrices. Note that $d(\mathrm{Eig}\ A, \mathrm{Eig}\ B) = \|(\mathrm{Eig}\ A, \mathrm{Eig}\ B)\|$. Since Weyl's inequality (1) can be generalized to the inequality (12) for Hermitian matrices, one is tempted to conjecture that the inequality (3) for unitary matrices can be similarly generalized. This, however, is not the case. It was proved by R. Bhatia, C. Davis and A. McIntosh (1983) that for A,B unitary

$$|||(\mathrm{Eig}\ A,\ \mathrm{Eig}\ B)||| \leq \frac{\pi}{2}\ |||A-B|||. \tag{14}$$

Further, no constant smaller than $\pi/2$ can replace it in this statement.

It is not known what the best such inequality for normal matrices would be. If A,B and A-B are all normal then it was shown by R. Bhatia (1982) that

$$|||(\mathrm{Eig}\ A,\ \mathrm{Eig}\ B)||| \leq |||A-B|||. \tag{15}$$

A more interesting result was obtained by Hoffman and Wielandt (1953). They showed that for any two normal matrices A,B and for the Frobenius norm we have

$$\|(\text{Eig } A, \text{ Eig } B)\|_F \leq \|A-B\|_F. \tag{16}$$

Several other results related to these may be found in the text.

This monograph is reasonably self-contained. In Chapter 1 we have collected some statements which are used in later chapters. These results may be found in standard books, references to which have been provided. The contents of Chapter 2 can also be found in the books of Schatten and of Gohberg and Krein. We have only picked up those results which provide an adequate working knowledge of unitarily-invariant norms and singular values. The contents of the remaining chapters are scattered in the research literature, though some of these results have already found a place in books. While organizing them into a coherent systematic whole, we have tried to find a common strand running through several results. Thus when two different proofs are available for the same result we have chosen not the one which is cleverer or the one which gives a little stronger conclusion but the one which leads to at least one more significant result. We, therefore, hope that the reader will not only find the inequalities interesting, but also the mathematics behind them.

1 Preliminaries

In this chapter we collect some miscellaneous facts for later use. Many of these are well known elementary statements. They are recalled here briefly and stated without proof.

§1. The marriage problem

Let $B = \{b_1,\ldots,b_n\}$ and $G = \{g_1,\ldots,g_n\}$ be two finite sets with the same cardinality and let R be a subset of $B\times G$. The triple (B,G,R) will be called a *society*. It will be called an *espousable society* if there is a bijection f from B to G whose graph is contained in R. The "marriage problem" is to decide when a society is espousable. (Think of B as a set of boys, G as a set of girls and say $(b_i, g_j) \in R$ iff the boy b_i and the girl g_j are compatible with each other. The problem is to decide whether it is possible to arrange a monogamous marriage in which each boy is married to a girl with whom he is compatible).

For each $i = 1,2,\ldots,n$ let $G_i = \{g_j : (b_i, g_j) \in R\}$, and for each k-tuple of indices $1 \leq i_1 < i_2 <\ldots< i_k \leq n$ let $G_{i_1\ldots i_k} = \bigcup_{r=1}^{k} G_{i_r}$. The nontrivial half of the following theorem is attributed to P. Hall.

THEOREM 1.1 (Hall): A society (B,G,R) is espousable iff for every $k = 1,2,\ldots,n$ and for every choice of indices $1 \leq i_1 <\ldots < i_k \leq n$

$$|G_{i_1\ldots i_k}| \geq k \tag{1.1}$$

Here, $|X|$ denotes the cardinality of the set X.

Notice that the condition (1.1) involves a gender asymmetry, in that, it only demands that for each set of boys we should be able to line up at least as many girls each of whom is compatible with one or more of these boys. If we banish such lack of reciprocity we get a stronger result.

For each $i = 1,2,\ldots,n$ let $B_i = \{b_j : (b_j, g_i) \in R\}$ and, as before, let

$B_{i_1 \ldots i_k} = \bigcup_{r=1}^{k} B_{i_r}$. For any real number t let {t} denote the smallest integer larger than t. The following theorem was proved by Elsner, Johnson, Ross and Schönheim.

THEOREM 1.2 : A society (B,G,R) is espousable iff for every $k = 1,2,\ldots,\{n/2\}$ and for every choice of indices $1 \le i_1 < \ldots < i_k \le n$

$$|G_{i_1 \ldots i_k}| \ge k \quad \underline{\text{and}} \quad |B_{i_1 \ldots i_k}| \ge k. \tag{1.2}$$

Proof : We shall show that this hypothesis implies that of Theorem 1.1. For any $r = 1,2,\ldots,n$ choose any indices $1 \le i_1 < \ldots < i_r \le n$. We have to show $|G_{i_1 \ldots i_r}| \ge r$. If $r \le \{n/2\}$ there is nothing to prove. Suppose there exist an $r \ge \{n/2\} + 1$ and indices $1 \le i_1 < \ldots < i_r \le n$ such that $|G_{i_1 \ldots i_r}| < r$. Then there are $n - r + 1$ girls, say $g_1, g_2, \ldots, g_{n-r+1}$ who are incompatible with each of the boys $b_{i_1}, \ldots, b_{i_r}$; i.e. $(b_{i_j}, g_m) \notin R$ for $j = 1,2,\ldots,r$ and $m = 1,2,\ldots,n-r+1$. So, $|B_{1 \ldots n-r+1}| \le n - r$. On the other hand, $n - r + 1 \le n - \{n/2\} \le \{n/2\}$ and so, by the hypothesis, $|B_{1 \ldots n-r+1}| \ge n - r + 1$. This contradiction proves our assertion. ■

We will use these theorems to estimate the distance between two unordered n-tuples of complex numbers under certain conditions.

Let $\{\lambda_1,\ldots,\lambda_n\}$ and $\{\mu_1,\ldots,\mu_n\}$ be two unordered n-tuples of complex numbers. (Some of the λ's or the μ's may be equal). Let L and M be the subsets of the complex plane which have the λ's and the μ's as their elements. (The cardinality of L or M may be less than n if the λ's or the μ's occur with multiplicity). The *Hausdorff distance* between the closed subsets L and M in the metric space $\mathbb{C}$ is defined, as usual, by

$$h(L,M) = \max(\nu(L,M), \nu(M,L))$$

where

$$\nu(L,M) = \sup_{\lambda\in L} \operatorname{dist}(\lambda,M)$$

The *optimal matching distance* between the n-tuples $\{\lambda_1,\ldots,\lambda_n\}$ and $\{\mu_1,\ldots,\mu_n\}$ is defined as

$$d(\{\lambda_1,\ldots,\lambda_n\}, \{\mu_1,\ldots,\mu_n\}) = \min_{\sigma\in S_n} \max_{1\leq i\leq n} |\lambda_i - \mu_{\sigma(i)}|,$$

where S_n denotes the group of permutations on n symbols.

The following statements are easy to verify.

1.3 $h(L,M) = \varepsilon$ iff ε is the smallest number with the property that each λ_i is within an ε-neighbourhood of some μ_j and *vice versa.*

1.4 (i) $h(L,M) \leq d(\{\lambda_1,\ldots,\lambda_n\}, \{\mu_1,\ldots,\mu_n\})$.

(ii) When $n = 2$, there is equality in the above inequality.

(iii) When $n = 3$, there may be strict inequality even in the absence of multiplicity.

(iv) The optimal matching distance may become arbitrarily large even when the Hausdorff distance remains bounded. This is best illustrated by the example

$$L = \{0,\ m-\varepsilon,\ m+\varepsilon\}, \qquad M = \{m, \varepsilon, -\varepsilon\}.$$

Later on, we will come across a situation where some additional information about the relative positions of $\{\lambda_1,\ldots,\lambda_n\}$, $\{\mu_1,\ldots,\mu_n\}$ is available. We will then need to use

THEOREM 1.5 : Let $\{\lambda_1,\ldots,\lambda_n\}$, $\{\mu_1,\ldots,\mu_n\}$ be two n-tuples of complex numbers. Let $\bar{D}(\lambda,\varepsilon)$ denote the closed disk with centre λ and radius ε. Let C denote any connected component of the set $\bigcup_{i=1}^{n} \bar{D}(\lambda_i,\varepsilon)$ or of the set $\bigcup_{i=1}^{n} \bar{D}(\mu_i,\varepsilon)$. If each such C contains as many λ's as it contains μ's then the optimal matching distance between these n-tuples is bounded by $n\varepsilon$ if n is odd and by $(n-1)\varepsilon$ if n is even.

Proof : Let B and G denote the sets $B = \{\lambda_1, \ldots, \lambda_n\}$, $G = \{\mu_1, \ldots, \mu_n\}$. (Here we allow ourselves a misuse of notation. Some of the λ's or the μ's may be identical. But for this argument we regard them as distinguishable, so that $|B| = |G| = n$). Say that λ_i and λ_j are joined by a p-string if there exists a sequence $\lambda_{i_1}, \lambda_{i_2}, \ldots, \lambda_{i_p}$ such that $\lambda_{i_1} = \lambda_i$, $\lambda_{i_p} = \lambda_j$ and $\overline{D}(\lambda_{i_k}, \varepsilon)$ has a nonempty intersection with $\overline{D}(\lambda_{i_{k+1}}, \varepsilon)$ for $k = 1,2,\ldots,p-1$. In the same way, define the joining of a μ_i with a μ_j by a p-string.

Note that if λ_i and λ_j are joined by a p-string then $|\lambda_i - \lambda_j| \leq 2(p-1)\varepsilon$.

Now define a relation $R \subset B \times G$ as follows. Say that $(\lambda_i, \lambda_j) \in R$ either if there exists a μ_k in $\overline{D}(\lambda_i, \varepsilon)$ such that μ_k and μ_j are joined by a p-string for some $p \leq \{n/2\}$, or if there exists a λ_m in $\overline{D}(\mu_j, \varepsilon)$ such that λ_m and λ_i are joined by a p-string for some $p \leq \{n/2\}$.

Note that if $(\lambda_i, \mu_j) \in R$ then $|\lambda_i - \mu_j| \leq \varepsilon + 2(\{n/2\} - 1)\varepsilon = (2\{n/2\} - 1)\varepsilon$. And this last quantity is equal to $(n-1)\varepsilon$ if n is even and $n\varepsilon$ if n is odd.

To prove the theorem it suffices to show that the society (B,G,R) is espousable. For this we use Theorem 1.2. Since there is a complete symmetry in the definition of R in our problem, we only need to verify that the first of the conditions (1.2) is satisfied for $k = 1,2,\ldots, \{n/2\}$.

First consider the case when each connected component C of the set $\bigcup_{i=1}^{n} \overline{D}(\mu_i, \varepsilon)$ is formed of no more than $\{n/2\}$ of these disks. By the hypothesis of the theorem all λ's must lie in one of these components. Each λ is then related to a μ if both of them lie in the same component and so $|G_{i_1 \ldots i_k}| \geq k$ for $k = 1,2,\ldots, \{n/2\}$. (In fact in this case this is true for all k).

Now suppose there is a connected component C of the set $\bigcup_{i=1}^{n} \overline{D}(\mu_i, \varepsilon)$ which is constituted of more than $\{n/2\}$ of these disks. Note that there can be only one such component. Any λ lying in this C is then related to $\{n/2\}$ of the μ's. Thus if $\lambda_{i_1}, \ldots, \lambda_{i_k}$ are chosen arbitrarily for any $k \leq \{n/2\}$ then those of this set which lie in C are all related to $\{n/2\}$ μ's, and for each λ not lying in C we can appeal to the earlier case. So, once again the condition $|G_{i_1 \ldots i_k}| \geq k$ is satisfied for $k = 1,2,\ldots, \{n/2\}$.

This proves the theorem. ∎

With a weaker hypothesis we get a weaker result:

THEOREM 1.6 : Let notations be as above. Suppose every connected component of the set $\bigcup_{i=1}^{n} \overline{D}(\lambda_i,\varepsilon)$ contains as many λ's as μ's. Then the optimal matching distance between these n-tuples is bounded by $(2n-1)\varepsilon$.

Proof : As in the earlier proof define a relation R by saying $(\lambda_i,\mu_j) \in R$ if $|\lambda_i-\mu_j| \leq (2n-1)\varepsilon$. Use Theorem 1.1. ∎

§2. Birkhoff's theorem

An nxn matrix A is called *doubly stochastic* if its entries a_{ij} satisfy the following conditions

(i) $a_{ij} \geq 0$ for $i,j = 1,2,\ldots,n$;

(ii) $\sum_{j=1}^{n} a_{ij} = 1$ for $i = 1,2,\ldots,n$;

(iii) $\sum_{i=1}^{n} a_{ij} = 1$ for $j = 1,2,\ldots,n$.

A matrix is called a *permutation matrix* if each row and each column have a single entry one and all other entries zero.

The following beautiful theorem also turns out to be extremely useful.

THEOREM 2.1 (Birkhoff) : The set Ω_n of nxn doubly stochastic matrices is convex; each permutation matrix is an extreme point of Ω_n; the set Ω_n is the convex hull of the set of permutation matrices.

§3. Majorization

Let $x = (x_1,\ldots,x_n)$ be an element of $\mathbb{R}^n$. Rearrange the components of x in decreasing order and denote them as

$$x_{[1]} \geq x_{[2]} \geq \cdots \geq x_{[n]}.$$

The same components arranged in increasing order will be denoted as

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}.$$

Let

$$x_{\downarrow} = (x_{[1]},\ldots,x_{[n]}),$$

$$x_{\uparrow} = (x_{(1)},\ldots,x_{(n)}).$$

Let $x,y \in \mathbb{R}^n$. We say that x is *majorized* by y if

$$\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]}, \quad k = 1,2,\ldots,n$$

and if equality holds in the above inequality for k = n. We will use the notation $x \prec y$ to say x is majorized by y.

The following theorem due to Hardy, Littlewood and Pólya is a basic result in the study of majorization.

<u>THEOREM 3.1 (The HLP Theorem)</u> : For $x,y \in \mathbb{R}^n$ the following conditions are equivalent

(i) $x \prec y$,

(ii) there is a doubly stochastic matrix A such that $x = Ay$,

(iii) the vector x lies in the convex hull of the n! vectors obtained by permuting the coordinates of the vector y.

We will also have occasion to use the related notion of (weak) submajorization. We say x is *(weakly) submajorized* by y if

$$\sum_{i=1}^{k} x_{[i]} \leq \sum_{i=1}^{k} y_{[i]}, \quad k = 1,2,\ldots,n.$$

This relation between x and y is denoted by $x \prec_w y$.

There is an analogue of the HLP Theorem which characterizes submajorizatior

An nxn matrix $A = ((a_{ij}))$ is called *doubly substochastic* if it satisfies the following conditions

(i) $a_{ij} \geq 0$ for $i,j = 1,2,\ldots,n$;

(ii) $\sum_{j=1}^{n} a_{ij} \leq 1$ for $i = 1,2,\ldots,n$;

(iii) $\sum_{i=1}^{n} a_{ij} \leq 1$ for $j = 1,2,\ldots,n$.

Let $\mathbb{R}^n_+$ denote the set of all vectors in $\mathbb{R}^n$ with nonnegative components. The following theorem runs parallel to the HLP Theorem.

THEOREM 3.2 : For $x,y \in \mathbb{R}^n_+$ the following conditions are equivalent

(i) $x \prec_w y$,

(ii) there is a doubly substochastic matrix A such that $x = Ay$,

(iii) the vector x lies in the convex hull of the $2^n\, n!$ vectors obtaine from y by permutations and sign changes of its coordinates, i.e. x lies in the convex hull of all vectors z which have the form

$$z = (\varepsilon_1\, y_{\sigma(1)}, \ldots, \varepsilon_n\, y_{\sigma(n)}),$$

where σ is a permutation and each $\varepsilon_j = \pm 1$.

If $x,y \in \mathbb{R}^n$ we say $x \leq y$ if $x_j \leq y_j$ for all $j = 1,2,\ldots,n$. A map $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ is called *monotone increasing* if $\Phi(x) \leq \Phi(y)$ whenever $x \leq y$. It is called *convex* if $\Phi(tx + (1-t)y) \leq t\,\Phi(x) + (1-t)\Phi(y)$ for $0 \leq t \leq 1$ and for all $x,y \in \mathbb{R}^n$.

A map $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ is called *isotone* if $\Phi(x) \prec_w \Phi(y)$ whenever $x \prec y$. Isotone maps are also called *Schur - convex*, though some authors use this latter term only when $m = 1$. The map Φ is called *strongly isotone* if $\Phi(x) \prec_w \Phi(y)$ whenever $x \prec_w y$.

THEOREM 3.3 : Let $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ be a convex map. Suppose for every permutation matrix A of order n there exists a permutation matrix $\tilde{A}$ of order m such that $\Phi(Ax) = \tilde{A}\Phi(x)$ for every $x \in \mathbb{R}^n$. Then Φ is isotone. If, in addition, Φ is monotone increasing then it is strongly isotone.

COROLLARY 3.4 : Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function. Then the map induced by f on $\mathbb{R}^n$ is isotone. If, in addition, f is monotone increasing then the induced map is strongly isotone.

In particular, note that this implies that if $x \prec y$ then $|x| \prec_w |y|$. (Here $|x|$ denotes the vector $(|x_1|,\ldots,|x_n|)$.

COROLLARY 3.5 : If $\Phi : \mathbb{R}^n \to \mathbb{R}$ is convex and permutation-invariant then Φ is isotone. If, in addition, Φ is monotone increasing then it is strongly isotone.

§4. Tensor products

Let H be a Hilbert space of dimension n with inner product $\langle .,. \rangle$. The k-fold *tensor power of H* will be denoted by $\otimes^k H$. The tensor product of vectors $x_1,\ldots,x_k$ in H is denoted by $x_1 \otimes \ldots \otimes x_k$. The space $\otimes^k H$ has dimension n^k. The inner product in this Hilbert space is defined by

$$\langle x_1 \otimes \ldots \otimes x_k, \; y_1 \otimes \ldots \otimes y_k \rangle = \prod_{i=1}^{k} \langle x_i, y_i \rangle . \tag{4.1}$$

For $1 \le k \le n$, let S_k be the symmetric group of degree k. Define a map P_k on $\otimes^k H$ by defining it on product vectors as

$$P_k(x_1 \otimes \ldots \otimes x_k) = \frac{1}{k!} \sum_{\sigma} \operatorname{sgn}(\sigma) x_{\sigma(1)} \otimes \ldots \otimes x_{\sigma(k)},$$

where the summation runs over $\sigma \in S_k$ and $\operatorname{sgn}(\sigma)$ denotes the signature of the permutation σ. Then P_k is an orthogonal projection and the range of this projection is denoted as $\Lambda^k H$. This space is called the k-fold *exterior power* or the *Grassman power* or the *antisymmetric tensor power* of H.

The exterior product of $x_1,\ldots,x_k$ is an element of $\Lambda^k H$ defined by

$$x_1 \wedge \ldots \wedge x_k = (k!)^{1/2} P_k(x_1 \otimes \ldots \otimes x_k).$$

The inner product in the space $\Lambda^k H$ is given by

$$\langle x_1 \wedge \ldots \wedge x_k, \; y_1 \wedge \ldots \wedge y_k \rangle = \det((\langle x_i, y_j \rangle)).$$

Here $((a_{ij}))$ denotes a matrix with entries a_{ij}.

The dimension of $\Lambda^k H$ is $\binom{n}{k}$.

If $e_1,\dots,e_n$ is an orthonormal basis for H then $e_{i_1}\otimes\dots\otimes e_{i_k}$, $1 \leq i_1 \leq\dots\leq i_k \leq n$, is an orthonormal basis for $\otimes^k H$ and $e_{i_1}\Lambda\dots\Lambda e_{i_k}$, $1 \leq i_1 <\dots< i_k \leq n$, is an orthonormal basis for $\Lambda^k H$.

Given a linear operator A on H we can define its kth tensor power $\otimes^k A$ as an operator on $\otimes^k H$ by defining its action on product vectors as

$$\otimes^k A\,(x_1\otimes\dots\otimes x_k) = Ax_1\otimes\dots\otimes Ax_k.$$

This operator leaves invariant the subspace $\Lambda^k H$ of $\otimes^k H$. Its restriction to $\Lambda^k H$ is denoted as $\Lambda^k A$. This operator acts on product vectors as

$$\Lambda^k A(x_1\Lambda\dots\Lambda x_k) = Ax_1\Lambda\dots\Lambda Ax_k.$$

The map $A \to \otimes^k A$ has the functorial properties

$$\otimes^k(AB) = \otimes^k A.\otimes^k B$$

$$\otimes^k(I) = I$$

$$(\otimes^k A)^* = \otimes^k(A^*).$$

These properties are shared by the map $A \to \Lambda^k A$.

If A has eigenvalues $\alpha_1,\dots,\alpha_n$ then $\otimes^k A$ has eigenvalues $\alpha_{i_1}\dots\alpha_{i_k}$, $1 \leq i_1 \leq\dots\leq i_k \leq n$ and $\Lambda^k A$ has eigenvalues $\alpha_{i_1}\dots\alpha_{i_k}$, $1 \leq i_1 <\dots< i_k \leq n$.

For $1 \leq k \leq n$ let

$$Q_{k,n} = \{\alpha : \alpha = (\alpha_1,\dots,\alpha_k),\ \ 1 \leq \alpha_1 <\dots< \alpha_k \leq n\}$$

be the collection of strictly increasing sequences of k integers chosen from $1,2,\dots,n$. An element α of $Q_{k,n}$ is called a k-index. The cardinality of $Q_{k,n}$ is $\binom{n}{k}$. We order the set $Q_{k,n}$ by the usual lexicographic ordering.

For two k-indices $\alpha = (\alpha_1,\ldots,\alpha_k)$ and $\beta = (\beta_1,\ldots,\beta_k)$ we denote by $A[\alpha|\beta]$ the submatrix of the matrix A constructed by picking the $\alpha_1,\ldots,\alpha_k$ rows and the $\beta_1,\ldots,\beta_k$ columns of A. In other words, if A is an nxn matrix with entries a_{ij} then $A[\alpha|\beta]$ is a kxk matrix whose (ij)th entry is $a_{\alpha_i\beta_j}$. The entries of the matrix $\Lambda^k A$ of order $\binom{n}{k}$ are conveniently indexed by pairs α,β chosen from $Q_{k,n}$. The (α,β)th entry of $\Lambda^k A$ is det $A[\alpha|\beta]$. Notice that the trace of $\Lambda^k A$ is the sum of the kxk principal minors of A.

Thus if the characteristic polynomial of the matrix A is written as

$$\chi_A(t) = t^n - a_1 t^{n-1} + a_2 t^{n-2} - \ldots + (-1)^n a_n,$$

then

$$a_k = \text{tr } \Lambda^k A, \quad k = 1,2,\ldots,n.$$

Notes and references for Chapter 1

Theorem 1.1, known as Hall's Marriage Theorem can be found in any text on combinatorial mathematics, e.g., Wilson [1]. It was first proved in Hall [1]. Theorem 1.2 was proved recently by Elsner, Johnson, Ross and Schönheim [1]. They actually prove a little more general graph theoretic theorem. The proof given here is adapted from there. Theorem 1.5 is proved in the above paper of Elsner et al. Theorem 1.6 has been known for a long time. It occurs in Ostrowski [1] and also in his book [3].

Doubly stochastic matrices have been studied extensively. Theorem 2.1 was proved by Birkhoff [1]. Several different proofs can be found in the literature now.

The idea of majorization has been used widely by analysts, physicists and economists. It forms the basis of the proof of several inequalities. See the classic work of Hardy, Littlewood and Pólya [1], the more recent comprehensive treatise of Marshall and Olkin [1] or the concise lecture notes of Ando [1]. The proofs of the theorems in section 3 and several related results can be found in these sources.

The material in section 4 can be found in any text on algebra, e.g. Lang [1]. More detailed treatment can be found in texts on multilinear algebra like Greub [1] or Marcus [1]. A quick but complete survey can be found in Marcus and Minc [1]. For a very brief summary with several interesting applications see Blokhuis and Seidel [1].

2 Singular values and norms

In this chapter we define a special class of norms, called the *unitarily-invariant norms*, on the space of linear operators. We prove two key theorems; one due to von Neumann shows that these norms arise as *symmetric gauge functions* of the singular values of an operator; another due to Ky Fan is a very useful tool in proving inequalities for this whole class of norms. We provide proofs of all statements (except those which can be found in any linear algebra text at the level of Halmos [1]).

§5. Singular values and polar decomposition

All linear operators, from now onwards, will be assumed to be acting on a fixed Hilbert space H of dimension n. As usual we will identify operators with matrices.

We call an operator A *positive* if the inner product $\langle Ax,x\rangle$ is nonnegative for all $x \in H$. A positive operator is Hermitian and all its eigenvalues are positive. A positive operator A has a unique positive square root (i.e. a positive operator B such that $B^2 = A$).

For any operator A, the operator A^*A is always positive. The positive square root of this operator will be denoted by $|A|$. The eigenvalues of $|A|$ are called the *singular values* of A. We will always count these singular values with multiplicity, we will always number them in decreasing order:

$$s_1(A) \geq s_2(A) \geq \ldots \geq s_n(A) \geq 0$$

and we will denote by Sing A the vector in $\mathbb{R}^n_+$ whose coordinates are the singular values of A.

The following facts are easily deduced from the polar decomposition theorem stated below.

THEOREM 5.1 (Polar decomposition) : Given a linear operator A, there exist a unitary operator U and a positive operator P such that A = UP. The operator P is unique, in fact, P = $|A|$; the operator U is unique if A is invertible.

THEOREM 5.2 (Singular value decomposition) : Given a linear operator A, there exist unitary operators U and V and a diagonal operator D with positive entries on the diagonal such that A = UDV. The diagonal entries of D are the singular values of A.

Note the above two theorems can be derived from each other using the spectral theorem.

THEOREM 5.3 : Let A,B be two linear operators. Then the following two conditions are equivalent

(i) Sing A = Sing B,
(ii) there exist unitary operators U,V such that A = UBV.

If M(n) denotes the space of nxn matrices and U(n) the multiplicative group of unitary matrices then we can think of $A \to UAV$ as an action of the group $U(n) \times U(n)$ on M(n). Theorem 5.3 says that Sing A is a complete invariant for this action.

Let A = UP be the polar decomposition of A. Choose an orthonormal basis $e_1,\dots,e_n$ for H consisting of eigenvectors of P corresponding to the eigenvalues $s_1,\dots,s_n$ respectively. Let $Ue_j = f_j$, $j = 1,2,\dots,n$. Then f_j also form an orthonormal basis for H. Note

$$Ae_j = s_j Ue_j = s_j f_j,$$

$$A^* f_j = PU^* f_j = s_j e_j,$$

$$A^* Ae_j = s_j^2 e_j,$$

$$AA^* f_j = s_j^2 f_j.$$

Thus f_j are the eigenvectors of AA^* corresponding to the eigenvalues s_j^2 (which are the same as the eigenvalues of A^*A).

We say that e_j is a *left singular vector* and f_j is a *right singular vector* of A corresponding to the singular value s_j.

§6. The minmax principle

Let A be a Hermitian operator with eigenvalues arranged in a decreasing order as

$$\lambda_{[1]}(A) \geq \lambda_{[2]}(A) \geq \ldots \geq \lambda_{[n]}(A).$$

The following elementary result (see, e.g., Halmos [1]) is extremely useful.

THEOREM 6.1 (The minmax principle) : For every Hermitian operator A we have

$$\lambda_{[j]}(A) = \max_{M:\dim M=j} \ \min_{x\in M, \|x\|=1} \langle Ax,x\rangle$$

$$= \min_{N:\dim N=n-j+1} \ \max_{x\in N, \|x\|=1} \langle Ax,x\rangle$$

Here M and N denote subspaces of the Hilbert space $\mathcal{H}$ (of dimension n) on which A is acting.

COROLLARY 6.2 (The minmax principle for singular values) : For any operator A we have

$$s_j(A) = \max_{M:\dim M=j} \ \min_{x\in M, \|x\|=1} \|Ax\|$$

$$= \min_{N:\dim N=n-j+1} \ \min_{x\in N, \|x\|=1} \|Ax\|$$

COROLLARY 6.3 : Let A and B be Hermitian operators such that $A \leq B$ (i.e. B-A is a positive operator). Then

$$\lambda_{[j]}(A) \leq \lambda_{[j]}(B), \qquad j = 1,2,\ldots,n.$$

We denote by $\|A\|$ the operator norm of A defined by

$$\|A\| = \sup_{\|x\|=1} \|Ax\|. \tag{6.1}$$

We have

$$\|A\| = s_1(A). \tag{6.2}$$

PROPOSITION 6.4 : Let A,B be any two operators. Then

$$s_j(BA) \leq \|B\| \; s_j(A),$$

$$s_j(AB) \leq \|B\| \; s_j(A).$$

Proof : We have, for every vector x,

$$\begin{aligned} \langle A^* B^* BA\, x,x\rangle &= \|BAx\|^2 \\ &\leq \|B\|^2 \; \|Ax\|^2 \\ &= \|B\|^2 \; \langle A^* Ax,x\rangle \; . \end{aligned}$$

So, we have the operator inequality

$$A^* B^* BA \leq \|B\|^2 \; A^* A.$$

So, by Corollary 6.3,

$$\begin{aligned} s_j^2(BA) &= \lambda_{[j]}(A^*B^*BA) \\ &\leq \|B\|^2 \lambda_{[j]}(A^*A) \\ &= \|B\|^2 s_j^2(A). \end{aligned}$$

This proves the first inequality. The second follows from this by taking adjoints. ∎

We next derive an extremal characterization of the sum of the k highest eigenvalues of a Hermitian operator. For this it will be convenient to use a multilinear device.

Given an operator A on H define an operator $A^{(k)}$ on $\otimes^k H$ by

$$A^{(k)} = A\otimes I\otimes\ldots\otimes I + I\otimes A\otimes\ldots\otimes I + \ldots + I\otimes I\otimes\ldots\otimes A$$

where, there are k summands on the right hand side. Let $A^{[k]}$ denote the restriction of $A^{(k)}$ to $\Lambda^k H$. Then

$$A^{[k]}(x_1\Lambda\ldots\Lambda x_k) = Ax_1\Lambda\ldots\Lambda x_k + x_1\Lambda Ax_2\Lambda\ldots\Lambda x_k + \ldots + x_1\Lambda x_2\Lambda\ldots\Lambda Ax_k. \tag{6.3}$$

So, if A has eigenvalues α_i with respective eigenvectors x_i, $1 \leq i \leq n$, then $A^{[k]}$ has eigenvalues $\alpha_{i_1} + \ldots + \alpha_{i_k}$ with respective eigenvectors $x_{i_1}\Lambda\ldots\Lambda x_{i_k}$, $1 \leq i_1 <\ldots< i_k \leq n$.

In particular, this means that if A is normal/Hermitian/positive then so is $A^{[k]}$.

<u>THEOREM 6.5 (Ky Fan's maximum principle)</u> : Let A be any Hermitian operator. Then for k = 1,2,...,n, we have

$$\sum_{j=1}^{k} \lambda_{[j]}(A) = \max \sum_{j=1}^{k} \langle Ax_j, x_j\rangle$$

where the maximum is taken over all orthonormal k-tuples $\{x_1,\ldots,x_k\}$ varying in H.

Proof : For brevity, put $\lambda_j = \lambda_{[j]}(A)$ for the duration of this proof. Let $e_1,\dots,e_n$ be orthonormal eigenvectors of A for the eigenvalues $\lambda_1,\dots,\lambda_n$. The highest eigenvalue of the Hermitian operator $A^{[k]}$ is $\lambda_1 +\dots+ \lambda_k$. Apply the minmax principle to this operator in the space $\Lambda^k H$, to get

$$\lambda_1 +\dots+ \lambda_k = \sup_{x\in\Lambda^k H, \|x\|=1} \langle A^{[k]} x,x\rangle$$

$$\geq \sup \langle A^{[k]}(x_1\Lambda\dots\Lambda x_k), \quad x_1\Lambda\dots\Lambda x_k\rangle$$

where the last supremum is taken over all orthonormal k-tuples $\{x_1,\dots,x_k\}$ chosen from H. Now note,

$$\langle A^{[k]}(x_1\Lambda\dots\Lambda x_k), \quad x_1\Lambda\dots\Lambda x_k\rangle = \langle Ax_1\Lambda\dots\Lambda x_k, \; x_1\Lambda\dots\Lambda x_k\rangle$$

$$+ \langle x_1\Lambda \; Ax_2\Lambda\dots\Lambda x_k, \; x_1\Lambda\dots\Lambda x_k\rangle$$

$$+ \dots$$

$$+ \langle x_1\Lambda\dots\Lambda Ax_k, \; x_1\Lambda\dots\Lambda x_k\rangle.$$

The first of the above terms is the determinant

$$\begin{vmatrix} \langle Ax_1,x_1\rangle & \langle Ax_1,x_2\rangle & \dots & \langle Ax_1,x_k\rangle \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{vmatrix}$$

and this is equal to $\langle Ax_1,x_1\rangle$. In the same way, the jth summand in the above relation equals $\langle Ax_j,x_j\rangle$. So,

$$\lambda_1 +\dots+ \lambda_k \geq \sup \sum_{j=1}^{k} \langle Ax_j,x_j\rangle .$$

Now choosing $x_j = e_j$ we find that this supremum is actually attained and is equal to $\lambda_1 + \ldots + \lambda_k$. ■

<u>COROLLARY 6.6</u> : Let A,B be any two Hermitian operators on H. Then for $1 \leq k \leq n$,

$$\sum_{j=1}^{k} \lambda_{[j]}(A+B) \leq \sum_{j=1}^{k} \lambda_{[j]}(A) + \sum_{j=1}^{k} \lambda_{[j]}(B).$$

<u>COROLLARY 6.7</u> : Let A,B be any two operators on H. Then, for $1 \leq k \leq n$,

$$\sum_{j=1}^{k} s_j(A+B) \leq \sum_{j=1}^{k} s_j(A) + \sum_{j=1}^{k} s_j(B).$$

<u>Proof</u> : This statement can be derived from that of Corollary 6.6 by a very useful device due to Wielandt. Note that if A is any matrix then the matrix

$$\tilde{A} = \begin{bmatrix} 0 & A \\ A^* & 0 \end{bmatrix}$$

is Hermitian and the eigenvalues of $\tilde{A}$ are the singular values of A together with their negatives.

Apply Corollary 6.6 to $\tilde{A}$ and $\tilde{B}$ to get Corollary 6.7. ■

<u>COROLLARY 6.8</u> : Let A be a Hermitian matrix. Let d and λ denote the vectors in $\mathbb{R}^n$ whose coordinates are the diagonal entries of A and the eigenvalues of A, respectively. Then d is majorized by λ.

<u>§7. Symmetric gauge functions and norms</u>

Let $|||\cdot|||$ be a norm on the space of $n \times n$ matrices. Such a norm is called *unitarily-invariant* if for all A and for any unitary U,V we have

$$|||A||| = |||UAV||| \tag{7.1}$$

We shall always assume that a norm is normalized so that the diagonal matrix with a single entry 1 and the other entries zero has norm 1. This normalization is inessential but convenient.

An example of such a norm is the *operator norm* $\|A\|$ defined by (6.1). Another example is the *Frobenius norm* or the *Hilbert-Schmidt norm* defined by

$$\|A\|_F = (\text{tr } A^*A)^{1/2} = \left(\sum_{i,j} |a_{ij}|^2 \right)^{1/2}, \tag{7.2}$$

where tr A denotes the trace of A and a_{ij} are the entries of the matrix A.

We shall adopt the following notational convention. The symbol $\|\cdot\|$ without a suffix will always denote the operator norm. With a suffix it will denote some other norm, as in (7.2) above. The symbol $|||\cdot|||$ will stand for any of the family of unitarily-invariant norms.

From Theorem 5.3 it is clear that a unitarily-invariant norm is a function only of the singular values of an operator. What kind of a function can it be? Once this question is answered, we will be able to construct several other such norms.

A map $\Phi : \mathbb{R}^n \to \mathbb{R}_+$ is called a *symmetric gauge function* if it satisfies the following conditions

(i) $\Phi(x) \geq 0$; $\Phi(x) = 0$ iff $x = 0$

(ii) $\Phi(\alpha x) = |\alpha| \, \Phi(x)$ for all $\alpha \in \mathbb{R}$

(iii) $\Phi(x+y) \leq \Phi(x) + \Phi(y)$

(iv) $\Phi(\Pi x) = \Phi(x)$ for every permutation matrix Π

(v) $\Phi(\varepsilon_1 x_1, \ldots, \varepsilon_n x_n) = \Phi(x_1, \ldots, x_n)$ for $\varepsilon_j = \pm 1$, $j = 1,2,\ldots,n$

(vi) $\Phi(1,0,\ldots,0) = 1$

(The last condition is an inessential normalization).

Examples of such functions are

$$\Phi_1(x) = \sum_{j=1}^{n} |x_j|$$

$$\Phi_\infty(x) = \max_{1 \leq j \leq n} |x_j|$$

$$\Phi_p(x) = \left(\sum_{j=1}^{n} |x_j|^p \right)^{1/p}, \quad 1 \le p \le \infty$$

$$\Phi_k(x) = \max_{1 \le i_1 < \ldots < i_k \le n} (|x_{i_1}| + \ldots + |x_{i_k}|), \quad 1 \le k \le n.$$

Note that Φ_p for p = 1 and Φ_k for k = 1 mean quite different things. To avoid confusion we will always say Φ_p for p = 1 and Φ_k for k = 1 for these objects, reserving the symbols p and k as above.

PROPOSITION 7.1 : Every symmetric gauge function Φ satisfies the following properties

(i) for every $x \in \mathbb{R}^n$ and for real numbers $t_1,\ldots,t_n$ between 0 and 1

$$\Phi(t_1x_1,\ldots,t_nx_n) \le \Phi(x_1,\ldots,x_n),$$

(ii) $\Phi_\infty(x) \le \Phi(x) \le \Phi_1(x)$ for all $x \in \mathbb{R}^n$,

(iii) Φ is continuous.

Proof : (i) By property (v) of symmetric gauge functions assume without loss of generality that $x_j \ge 0$ for all j. Note that it is enough to prove the statement for the case when only one $t_j \ne 1$. We have

$$\Phi(x_1,\ldots,tx_j,\ldots,x_n)$$

$$= \Phi\left(\frac{1+t}{2}x_1 + \frac{1-t}{2}x_1,\ldots, \frac{1+t}{2}x_j + \frac{1-t}{2}(-x_j),\ldots, \frac{1+t}{2}x_n + \frac{1-t}{2}x_n\right)$$

$$\le \frac{1+t}{2}\Phi(x_1,\ldots,x_n) + \frac{1-t}{2}\Phi(x_1,\ldots,-x_j,\ldots,x_n)$$

$$= \Phi(x_1,\ldots,x_n).$$

This proves (i). The other two assertions are equally easy to prove. ∎

COROLLARY 7.2 : Every symmetric gauge function is monotonically increasing on $\mathbb{R}^n_+$.

THEOREM 7.3 : For $x,y \in \mathbb{R}^n_+$ the following two conditions are equivalent

(i) $x \prec_w y$

(ii) $\Phi(x) \leq \Phi(y)$ for every symmetric gauge function Φ.

Proof : Suppose (ii) holds. Choose, in particular, the symmetric gauge functions Φ_k, $1 \leq k \leq n$. By definition, these inequalities mean $x \prec_w y$.

Conversely, note that any symmetric gauge function is convex, permutation invariant and monotone increasing on $\mathbb{R}^+_n$. So, by Corollary 3.5 Φ is strongl isotone. Thus (i) implies (ii). ■

Remark : The proof could also have been based on Theorem 3.2. By that theorem $x \prec_w y$ implies that x is a convex combination of vectors obtained from y by permutations and sign changes of coordinates. On each of these vectors Φ takes the same value $\Phi(y)$. This together with properties (ii) and (iii) of symmetric gauge functions implies $\Phi(x) \leq \Phi(y)$.

THEOREM 7.4 : Let Φ be a symmetric gauge function on $\mathbb{R}^n$. For $A \in M(n)$, the space of nxn matrices, define

$$|||A|||_\Phi = \Phi(s_1(A),\dots,s_n(A)).$$

Then this defines a unitarily-invariant norm.

Conversely given any unitarily-invariant norm $|||\cdot|||$ on M(n) define a function on $\mathbb{R}^n_+$ by

$$\Phi_{|||\cdot|||}(s_1,\dots,s_n) = |||A|||$$

where A is any operator with Sing A = $(s_1,\dots,s_n)$. Then this defines a symmetric gauge function.

Proof : Since Sing A = Sing UAV, clearly $|||\cdot|||_\Phi$ is unitarily-invariant. To show that it is a norm we need to show that it satisfies the triangle inequality, the other properties of a norm being obviously satisfied. For this use Corollary 6.7, which implies

$$\text{Sing }(A+B) \prec_w \text{Sing } A + \text{Sing } B.$$

So, by Theorem 7.3,

$$\Phi(\text{Sing }(A+B)) \leq \Phi(\text{Sing } A + \text{Sing } B)$$

$$\leq \Phi(\text{Sing } A) + \Phi(\text{Sing } B)$$

Thus,

$$|||A+B|||_\Phi \leq |||A|||_\Phi + |||B|||_\Phi .$$

The proof of the second part of the Theorem is left to the reader. ■

Thus there is a one-to-one correspondence between unitarily-invariant norms on M(n) and symmetric gauge functions on $\mathbb{R}^n$.

Two important families of unitarily-invariant norms are the *Schatten p norms* defined for $1 \leq p \leq \infty$ by

$$\|A\|_p = \left(\sum_{j=1}^{n} (s_j(A))^p \right)^{1/p}$$

and the *Ky Fan k norms* defined for k = 1,2,...,n by

$$\|A\|_k = \sum_{j=1}^{k} s_j(A).$$

Note that the Schatten p norm for $p = \infty$ and the Ky Fan k norm for k = 1 coincide with the *operator norm*. The Schatten p norm for p = 2 is the

Frobenius norm. The Schatten p norm for p = 1 is equal to the Ky Fan k norm for k = n. It is called the *trace norm* and is denoted by $\|A\|_{tr}$.

THEOREM 7.5 : Let A,B be two linear operators. If $\|A\|_k \leq \|B\|_k$ for all Ky Fan norms k = 1,2,...,n, then $|||A||| \leq |||B|||$ for every unitarily-invariant norm.

Proof : Use Theorems 7.3 and 7.4. ■

PROPOSITION 7.6 : Every unitarily-invariant norm dominates the operator norm and is dominated by the trace norm, i.e.,

$$\|A\| \leq |||A||| \leq \|A\|_{tr}$$

for all operators A.

Proof : Use Proposition 7.1 (ii). ■

PROPOSITION 7.7 : For any three operators A,B and C we have

$$|||BAC||| \leq \|B\| \ |||A||| \ \|C\| \tag{7.3}$$

Proof : By Proposition 6.4

$$s_j(BAC) \leq \|B\| \ \|C\| \ s_j(A).$$

So, by Theorem 7.3

$$\Phi(\text{Sing } BAC) \leq \|B\| \ \|C\| \ \Phi(\text{Sing } A).$$

So 7.3 follows from Theorem 7.4 now. ■

A norm ν on $M(n)$ is called a *symmetric norm* if for all A,B,C

$$\nu(BAC) \leq \|B\| \; \nu(A) \; \|C\|.$$

Proposition 7.7 says that every unitarily-invariant norm is symmetric. Conversely, note that if ν is a symmetric norm and if U,V are unitary then

$$\nu(UAV) \leq \|U\| \; \nu(A) \; \|V\| = \nu(A)$$

and

$$\nu(A) = \nu(U^{-1}UAV\,V^{-1}) \leq \nu(UAV).$$

So ν is unitarily-invariant.

In Chapter 4 we will need another property of unitarily-invariant norms. Let $P_1,P_2,\ldots,P_r$ be a complete family of mutually orthogonal projection operators in H. The *pinching* of an operator A by the projections $P_1,\ldots,P_r$ is the operator

$$C(A) = \sum_{i=1}^{r} P_i \; A \; P_i. \tag{7.4}$$

In an appropriate coordinate system the pinching operation takes the matrix A to a block diagonal matrix consisting of r diagonal blocks whose sizes are the ranks of the projections P_i. This matrix is obtained from A by replacing the entries outside these blocks by zeros.

A pinching induced by a family of r projections as in 7.4 will be called an r-pinching. For $j = 1,2,\ldots,r-1$, put $Q_j = P_1 + \ldots + P_j$ and define a 2-pinching C_j by

$$C_j(A) = Q_j \; A \; Q_j + (I-Q_j) \; A \; (I-Q_j).$$

It is easy to see that the pinching 7.4 can be expressed as

$$C(A) = C_{r-1} \cdots C_2 \; C_1(A). \tag{7.5}$$

Thus an r-pinching can be obtained by successively applying 2-pinchings.

THEOREM 7.8 : Every unitarily-invariant norm is diminished by a pinching, i.e.,

$$|||C(A)||| \leq |||A|||$$

for every pinching C and for every operator A.

Proof : Because of the decomposition 7.5, it is enough to prove this when C is a 2-pinching. In a suitable coordinate system we can write in this case

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \qquad C(A) = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}.$$

But then we can write

$$C(A) = \frac{1}{2}(A + UAU^*)$$

where,

$$U = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix}.$$

Since U is a unitary operator the theorem follows. ■

Notes and references for Chapter 2

The minmax principle is a powerful tool in the study of variational properties of eigenvalues. Its power stems from the fact that it provides information about eigenvalues without any reference to eigenvectors or the characteristic polynomial. The principle was first stated by Fischer [1]. It was generalized and extended to wide classes of infinite-dimensional operators by Courant [1]. An early effective use for problems arising in differential equations was made by Courant [1] and by Weyl [1].

Consequently, it is referred to as the Fischer Minmax Theorem (as in Beckenbach and Bellman [1]) or the Courant-Fischer Minmax Theorem (as in Bellman [1] and Ando [1]). Parlett [1] traces its history back to Poincaré.

Singular values or "s-numbers" seem to have been introduced by E. Schmidt. Despite their usefulness in operator theory (see, for example, Gohberg and Krein [1]), in statistics (see, for example, Rao [1]), and in numerical analysis (see, for example, Stewart [1]), they have been benignly neglected by the authors of most linear algebra texts.

Weyl [2] initiated a study of inequalities for singular values and eigenvalues which was continued, among others, by Pólya, Horn and Fan. A complete, systematic and readable account of the work of these authors may be found in Gohberg and Krein [1].

Theorem 6.5 occurs in Ky Fan [1] with a different proof. (The minmax principle was generalized further by Wielandt; this is presented in the next chapter). The statement of Corollary 6.7 occurs in Fan [3]. The idea of using the matrix $\tilde{A}$ is attributed to Wielandt by Fan and Hoffman [1]. The majorization assertion of Corollary 6.8 is a famous result of Schur [1]. Though Schur proved this only for positive definite matrices, his proof works for all Hermitian matrices. This was observed by Mirsky [2]. It is easy to see that the statements of Theorem 6.5 and Corollary 6.8 are equivalent.

The connection between symmetric gauge functions and unitarily-invariant norms was pointed out by von Neumann [1]. An exhaustive account is given in the famous monograph of Schatten [1], whereas Mirsky [1] provides a quick complete introduction. Theorem 7.4 is due to von Neumann. Theorems 7.3 and 7.5 are due to Fan [3].

The pinching operator is called the diagonal-cell operator by Gohberg and Krein [1]. Theorem 7.8 is proved there (p.52) for the class of Ky Fan norms. That gives a different proof of Theorem 7.8. Davis [1] has studied several properties of pinchings, the effect of pinchings on eigenvalues, etc.

The operator norm is familiar to everyone with a first course in functional analysis. The other norms are less familiar, though they are used extensively in numerical analysis of matrices. There are two reasons for their usefulness. The operator norm is not easy to compute, whereas, the Frobenius norm, for instance, can easily be computed from the entries of a matrix. Second, to say that an operator has a small operator norm means that it is uniformly small. To quote Davis and Kahan [1] : "Saying an operator is smaller than ε everywhere is good if you can do it, but it may be more important and/or more feasible to say that it is smaller than $\varepsilon/10$ except on a subspace of small dimensionality. This sort of assertion involves the other unitary-invariant norms".

Extensive use of matrix norms was made by Householder [1] in analysing matrix processes. See also the book by Collatz [1]. Though the unitarily invariant norms are geometrically appealing, we must point out that there are important norms used by the numerical analyst which are not unitarily-invariant. For example, the norms

$$\|A\|_r = \max_i \sum_j |a_{ij}|$$

$$\|A\|_c = \max_j \sum_i |a_{ij}|$$

$$\|A\|_m = \max_{i,j} |a_{ij}|$$

are not unitarily-invariant.

Unitary invariance in the sense used here implies the weaker invariance property $|||A||| = |||UAU^{-1}|||$ but is not equivalent to it. For example, the norm $\nu(A) = \|A\| + |\text{tr } A|$ is invariant under the conjugate action of $U(n)$, $(A \to UAU^{-1})$, but is not unitarily-invariant in the above sense. The theory of such "weakly unitarily-invariant norms" is not as well developed as the Schatten-von Naumann Theory of unitarily-invariant norms. Several interesting results on such norms may be found in Fong and Holbrook [1] and in the references cited therein.

3 Spectral variation of Hermitian matrices

The spectral variation problem for the class of Hermitian matrices has been completely solved in the following sense. For any two Hermitian matrices a tight upper bound for the distance between their eigenvalues is known. Such bounds are known when the distance is measured in any unitarily-invariant norm. Further, in this case lower bounds for spectral variation are also known. All these results are presented in this chapter.

§8. Weyl's inequalities

The prototype of the spectral variation inequalities, to which most of this monograph is devoted, is the following result of H. Weyl:

THEOREM 8.1 : Let A and B be Hermitian matrices with eigenvalues $\lambda_{[1]}(A) \geq \ldots \geq \lambda_{[n]}(A)$ and $\lambda_{[1]}(B) \geq \ldots \geq \lambda_{[n]}(B)$ respectively. Then

$$\max_j |\lambda_{[j]}(A) - \lambda_{[j]}(B)| \leq \|A-B\|. \tag{8.1}$$

The proof of the Theorem can be based on

THEOREM 8.2 : For A,B Hermitian

$$\lambda_{[j]}(A) + \lambda_{[n]}(B) \leq \lambda_{[j]}(A+B) \leq \lambda_{[j]}(A) + \lambda_{[1]}(B). \tag{8.2}$$

Proof : Let $x_1,\ldots,x_n$ be eigenvectors of A for the eigenvalues $\lambda_{[1]}(A),\ldots,\lambda_{[n]}(A)$ respectively. Let M be the subspace spanned by $x_1,\ldots,x_j$. By the minmax principle (Theorem 6.1) we have

$$\lambda_{[j]}(A+B) \geq \min_{x\in M, \|x\|=1} \langle (A+B)x,x \rangle$$

$$\geq \min_{x\in M, \|x\|=1} \langle Ax,x \rangle + \min_{x\in M, \|x\|=1} \langle Bx,x \rangle$$

$$= \lambda_{[j]}(A) + \min_{x\in M, \|x\|=1} \langle Bx,x \rangle$$

$$\geq \lambda_{[j]}(A) + \min_{\|x\|=1} \langle Bx,x \rangle$$

$$= \lambda_{[j]}(A) + \lambda_{[n]}(B).$$

This proves the first inequality. To prove the second, write $A = (A+B) + (-B$ and use the above to get

$$\lambda_{[j]}(A) \geq \lambda_{[j]}(A+B) + \lambda_{[n]}(-B)$$

$$= \lambda_{[j]}(A+B) - \lambda_{[1]}(B) \quad \blacksquare$$

Proof of Theorem 8.1 : From the inequalities (8.2), we get

$$|\lambda_{[j]}(A+B) - \lambda_{[j]}(A)| \leq \max(|\lambda_{[1]}(B)|, |\lambda_{[n]}(B)|) = \|B\|.$$

By a change of variables this leads to (8.1). $\blacksquare$

REMARK 8.3 : We can prove Theorem 8.1 without recourse to the minmax principle. We have given the minmax proof because a generalization of Theorem 8.1 will soon be obtained by using a generalization of the minmax principle. We give another proof because this other idea will also be used at a later point.

Recall that for any operator A, the *numerical range* of A is the convex set

$$W(A) = \{\langle Ax,x\rangle \ : \ x \in H, \ \|x\| = 1\}. \qquad (8.3)$$

If A is Hermitian then

$$W(A) = [\lambda_{[n]}(A), \ \lambda_{[1]}(A)]. \qquad (8.4)$$

(In fact, we have used a part of this fact already in the proof of Theorem 8.2).

Let x_j be the eigenvector of A corresponding to the eigenvalue $\lambda_{[j]}(A)$ and y_j the eigenvector of B for the eigenvalue $\lambda_{[j]}(B)$. Let M be the linear span of $x_1,\ldots,x_j$ and N the linear span of $y_j,\ldots,y_n$. Then M and N have a nontrivial intersection. Choose a vector x in this intersection. Then $\langle Ax,x\rangle$ lies in the interval $[\lambda_{[j]}(A), \ \lambda_{[1]}(A)]$ and also in the interval $[\lambda_{[n]}(B), \ \lambda_{[j]}(B)]$. We have

$$\begin{aligned}\|A-B\| &= \sup_{\|v\|=1} |\langle (A-B)v,v\rangle| \\ &\geq |\langle Ax,x\rangle - \langle Bx,x\rangle| \\ &\geq \langle Ax,x\rangle - \langle Bx,x\rangle \\ &\geq \lambda_{[j]}(A) - \lambda_{[j]}(B).\end{aligned}$$

So, by symmetry

$$|\lambda_{[j]}(A) - \lambda_{[j]}(B)| \leq \|A-B\|.$$

Theorem 8.2 is a special case of a whole family of inequalities which follow from the minmax principle. Without aiming at completeness we record two such results.

In accordance with our notation in section 3, we denote the eigenvalues of A arranged in ascending order as

$$\lambda_{(1)}(A) \leq \dots \leq \lambda_{(n)}(A).$$

THEOREM 8.4 : Let A,B be any Hermitian operators on an n-dimensional space H. Then for any indices i,j satisfying $1 \leq i+j-1 \leq n$ we have

$$\lambda_{[i+j-1]}(A+B) \leq \lambda_{[i]}(A) + \lambda_{[j]}(B), \tag{8.5}$$

$$\lambda_{(i)}(A) + \lambda_{(j)}(B) \leq \lambda_{(i+j-1)}(A+B). \tag{8.6}$$

Proof : As before, let x_j be the eigenvector of A for the eigenvalue $\lambda_{[j]}(A)$ and y_j the eigenvector of B for the eigenvalue $\lambda_{[j]}(B)$. Let M be the linear span of $x_1,\dots,x_{i-1}$, N the linear span of $y_1,\dots,y_{j-1}$ and S the linear span of M and N. Then dim M = i-1, dim N = j-1 and dim $S \leq i+j-2$. Let k = dim S+1. Since $i+j-1 \geq k$, the ordering of the eigenvalues and the minmax principle imply

$$\begin{aligned}\lambda_{[i+j-1]}(A+B) &\leq \lambda_{[k]}(A+B) \\ &\leq \max_{x\in S^{\perp},\|x\|=1} \langle (A+B)x,x\rangle \\ &\leq \max_{x\in S^{\perp},\|x\|=1} \langle Ax,x\rangle + \max_{x\in S^{\perp},\|x\|=1} \langle Bx,x\rangle \\ &\leq \max_{x\in M^{\perp},\|x\|=1} \langle Ax,x\rangle + \max_{x\in N^{\perp},\|x\|} \langle Bx,x\rangle \\ &= \lambda_{[i]}(A) + \lambda_{[j]}(B).\end{aligned}$$

Here, $S^{\perp}$ means the orthogonal complement of S in H. This proves (8.5). The inequality (8.6) can be obtained from this by taking the negatives of all the operators involved. ■

Using either Theorem 8.4 or the argument used in Remark 8.3, one also gets

$$\|A-B\| \leq \max (|\lambda_{[1]}(A) - \lambda_{[n]}(B)|, |\lambda_{[n]}(A) - \lambda_{[1]}(B)|). \tag{8.7}$$

Let us combine (8.1) and (8.7) in a form which will be useful for later generalizations. For a given matrix A let Eig (A) denote the unordered n-tuple consisting of the eigenvalues of A. Let $\text{Eig}_{\downarrow}(A) = (\lambda_{[1]}(A),\ldots,\lambda_{[n]}(A))$ and $\text{Eig}_{\uparrow}(A) = (\lambda_{(1)}(A),\ldots,\lambda_{(n)}(A))$ be the decreasing and the increasing rearrangements of Eig (A). Denote also by $\text{Eig}_{\downarrow}(A)$ the diagonal matrix with entries $\lambda_{[1]}(A),\ldots,\lambda_{[n]}(A)$ down the diagonal. Other n-tuples will also describe diagonal matrices in the same fashion. With these notations we have

THEOREM 8.5 : For any two Hermitian matrices A and B we have

$$\|\text{Eig}_{\downarrow}(A) - \text{Eig}_{\downarrow}(B)\| \leq \|A-B\| \leq \|\text{Eig}_{\downarrow}(A) - \text{Eig}_{\uparrow}(B)\| \tag{8.8}$$

Proof : The norm of a diagonal matrix is the maximum modulus of its entries. This then is just a restatement of (8.1) and (8.7). ■

Note that both inequalities in (8.8) can become equalities, as is seen by choosing A,B to be appropriate diagonal matrices.

§9. The Lidskii-Wielandt theorem

The Courant-Fischer minmax principle is an extremal characterization of the eigenvalues $\lambda_{[j]}(A)$; Ky Fan's minmax principle (a part of which is stated as Theorem 6.5 is an extremal characterization of the sum of the top k eigenvalues of A. A generalization due to Wielandt subsumes these principles by giving an extremal characterization of the sum of any k eigenvalues of A. Though our main interest will be in some corollaries of this principle, which can be derived by other means as well, we give a proof of it here because it is the culmination of one circle of ideas studied in this monograph.

<u>THEOREM 9.1 (Wielandt's minmax principle)</u> : Let A be a Hermitian operator on the n-dimensional space $\mathcal{H}$. Then for any indices $1 \leq i_1 < \ldots < i_k \leq n$ we have

$$\lambda_{[i_1]}(A) + \ldots + \lambda_{[i_k]}(A)$$

$$= \max_{\substack{M_1 \subset \ldots \subset M_k \\ \dim M_j = i_j}} \quad \min_{\substack{x_j \in M_j \\ x_j \text{ orthonormal}}} \quad \sum_{j=1}^{k} \langle Ax_j, x_j \rangle$$

$$= \min_{\substack{N_1 \supset \ldots \supset N_k \\ \dim N_j = n-i_j+1}} \quad \max_{\substack{x_j \in N_j \\ x_j \text{ orthonormal}}} \quad \sum_{j=1}^{k} \langle Ax_j, x_j \rangle .$$

The proof is rather intricate and we split it into smaller propositions.

We will denote the subspace spanned by the vectors $x_1, \ldots, x_k$ as $[x_1, \ldots, x_k]$.

<u>LEMMA 9.2</u> : Let $W_1 \supset W_2 \supset \ldots \supset W_k$ be subspaces of $\mathcal{H}$ such that $\dim W_j \geq k-j+1$, $j = 1,2,\ldots,k$. Let $w_1, \ldots, w_{k-1}$ be linearly independent vectors such that $w_j \in W_j$ and let $U = [w_1, \ldots, w_{k-1}]$. Then there exists a nonzero vector u in $W_1 - U$ such that for the space $U + [u]$ we can find a basis $v_1, \ldots, v_k$ where $v_j \in W_j$.

<u>Proof</u> : This will be proved by induction on k. The statement is trivial when there is only one subspace involved. Assume it is true for k-1 subspaces.

Let $w_1, \ldots, w_{k-1}$ be given as above. Put $S = [w_2, \ldots, w_{k-1}]$ and apply the induction hypothesis to the subspaces $W_2 \supset \ldots \supset W_k$ to pick a vector v in

$W_2 - S$ such that $S + [v] = [v_2,\dots,v_k]$ for some linearly independent vectors $v_j \in W_j$, $j = 2,\dots,k$.

Suppose $v \in U$. Then $S + [v] = U$. Now U is a proper subspace of W_1 (because $\dim U = k-1$ and $\dim W_1 \geq k$). Choose a nonzero vector u in $W_1 - U$. Then $u, v_2,\dots,v_k$ form a basis for $[u] + U$ where $u \in W_1$ and $v_j \in W_j$ for $j = 2,\dots,k$.

Suppose $v \notin U$. Then $w_1 \notin S + [v]$, for if w_1 were a linear combination of $w_2,\dots,w_{k-1}$ and v, then v would be a linear combination of $w_1,\dots,w_{k-1}$, and hence be an element of U. So in this case $w_1, v_2,\dots,v_k$ span a k-dimensional space which coincides with $U + [v]$. Again $w_1 \in W_1$ and $v_j \in W_j$, $j = 2,\dots,k$. ■

<u>PROPOSITION 9.3</u> : Let $V_1 \subset V_2 \subset\dots\subset V_k$ be subspaces of H with $\dim V_j = i_j$, $1 \leq i_1 <\dots< i_k \leq n$. Let $W_1 \supset W_2 \supset\dots\supset W_k$ be subspaces of H with

$$\dim W_j = n - i_j + 1 = \operatorname{codim} V_j + 1.$$

Then there exist linearly independent vectors $v_j \in V_j$ and $w_j \in W_j$, $j = 1,2,\dots,k$ such that

$$[v_1,\dots,v_k] = [w_1,\dots,w_k].$$

<u>Proof</u> : We will apply induction on k. The statement is trivially true when there is only one V_1 and W_1. Assume it is true when $k-1$ pairs of subspaces are given.

Let $V_1 \subset\dots\subset V_k$ and $W_1 \supset\dots\supset W_k$ be given. By the induction hypothesis choose $v_j \in V_j$ and $w_j \in W_j$ for $j = 1,2,\dots,k-1$ such that $[v_1,\dots,v_{k-1}] = [w_1,\dots,w_{k-1}] = U$, say. Note U is a subspace of V_k.

Let $S_j = W_j \cap V_k$, $j = 1,\dots,k$.
We have,

$$\dim W_j + \dim V_k - \dim S_j \leq n.$$

So,

$$\dim S_j \geq i_k - i_j + 1 \geq k-j+1$$

Note that $S_1 \supset \ldots \supset S_k$ are subspaces of V_k and $w_j \in S_j$. Apply Lemma 9.2. There exists a vector u in $S_1 - U$ such that for the space $U + [u]$ we can find a basis $u_1,\ldots,u_k$ where $u_j \in S_j \subset W_j$, $j = 1,2,\ldots,k$. But $U + [u]$ is also spanned by $v_1,\ldots,v_{k-1}$, u where $u \in V_k$. Put $u = v_k$. We have thus found v_j and u_j in V_j and W_j respectively, for $j = 1,\ldots,k$, such that they span the same k-dimensional space. ■

Remark : By applying the Gram-Schmidt Process we could orthonormalize the vectors chosen above while they continue to satisfy the same properties.

Proof of Theorem 9.1 : We will prove the first statement (the second can be obtained from the first). Let $v_1,\ldots,v_n$ be eigenvectors of A for the eigenvalues $\lambda_{[1]}(A),\ldots,\lambda_{[n]}(A)$ respectively. Let

$$V_j = [v_1,v_2,\ldots,v_{i_j}], \quad j = 1,2,\ldots,k.$$

Then $V_1 \subset V_2 \subset \ldots \subset V_k$ and $\dim V_j = i_j$. Let $x_1,\ldots,x_k$ be any set of orthonormal vectors such that $x_j \in V_j$, $j = 1,2,\ldots,k$. Then $\langle Ax_j,x_j\rangle$ is larger than $\lambda_{[i_j]}(A)$. Since x_j were arbitrarily chosen we have shown

$$\inf_{\substack{x_j \in V_j \\ x_j \text{ orthonormal}}} \sum_{j=1}^{k} \langle Ax_j,x_j\rangle \geq \sum_{j=1}^{k} \lambda_{[i_j]}(A).$$

This infimum is actually attained for $x_j = v_{i_j}$.

We will be through if we show that for any subspaces $M_1 \subset \ldots \subset M_k$ with $\dim M_j = i_j$, we can find $x_j \in M_j$ such that

$$\sum_{j=1}^{k} \langle Ax_j, x_j \rangle \leq \sum_{j=1}^{k} \lambda_{[i_j]}(A).$$

With v_j as above put

$$W_j = [v_{i_j}, \ldots, n] \qquad j = 1,2,\ldots,k.$$

Then $W_1 \supset W_2 \supset \ldots \supset W_k$ and

$$\dim W_j = n - i_j + 1 = \operatorname{codim} M_j + 1.$$

So, by Proposition 9.3, there exist $x_j \in M_j$, $y_j \in W_j$, $j = 1,2,\ldots,k$ such that

$$[x_1, \ldots, x_k] = [y_1, \ldots, y_k] = W, \text{ say.}$$

As remarked, these vectors can actually be chosen to be orthonormal.

Let A_W denote the compression of A to the subspace W, i.e. A_W is the operator defined on W by $A_W x = P_W Ax$ where P_W is the projection onto W. Then A_W is a Hermitian operator on W. Note $\langle A_W x, x \rangle = \langle Ax, x \rangle$ for all $x \in W$. By the minmax principle

$$\lambda_{[j]}(A_W) = \min_{\substack{N \subset W \\ \dim N = k-j+1}} \max_{\substack{x \in N \\ \|x\|=1}} \langle A_W x, x \rangle$$

$$= \min_{\substack{N \subset W \\ \dim N = k-j+1}} \max_{\substack{x \in N \\ \|x\|=1}} \langle Ax, x \rangle$$

$$\leq \max_{\substack{x \in L \\ \|x\|=1}} \langle Ax, x \rangle ,$$

where, L is the orthogonal complement of $[y_1, \ldots, y_{j-1}]$ in W. Note that $L \subset W_j$. So,

$$\lambda_{[j]}(A_W) \leq \max_{x \in W_j, \|x\|=1} \langle Ax, x \rangle$$
$$= \lambda_{[i_j]}(A).$$

So, we have

$$\sum_{j=1}^{k} \langle Ax_j, x_j \rangle = \sum_{j=1}^{k} \langle A_W x_j, x_j \rangle = \operatorname{tr} A_W$$

$$= \sum_{j=1}^{k} \lambda_{[j]}(A_W) \le \sum_{j=1}^{k} \lambda_{[i_j]}(A).$$

This completes the proof. ■

Remark : Note that

$$\lambda_{[i_1]}(A) + \ldots + \lambda_{[i_k]}(A) = \sum_{j=1}^{k} \langle Av_{i_j}, v_{i_j} \rangle$$

where v_i are the eigenvectors of A. We have shown that the maximum in the first assertion of Wielandt's minmax principle is attained when $M_j = V_j = [v_1, v_2, \ldots, v_{i_j}]$, $j = 1,2,\ldots,k$. With this choice the minimum is attained for $x_j = v_{i_j}$, $j = 1,2,\ldots,k$.

As a corollary we have

THEOREM 9.4 (Lidskii's theorem) : Let A,B be Hermitian operators on the n-dimensional space $\mathcal{H}$. Then for any indices $1 \le i_1 < \ldots < i_k \le n$ we have

$$\sum_{j=1}^{k} \lambda_{[i_j]}(A+B) \le \sum_{j=1}^{k} \lambda_{[i_j]}(A) + \sum_{j=1}^{k} \lambda_{[j]}(B).$$

Proof : Use Theorem 9.1 to choose subspaces $M_1 \subset \ldots \subset M_k$ with $\dim M_j = i_j$ such that

$$\sum_{j=1}^{k} \lambda_{[i_j]}(A+B) = \min_{\substack{x_j \in M_j \\ x_j \text{ orthonormal}}} \sum_{j=1}^{k} \langle (A+B)x_j, x_j \rangle.$$

By Ky Fan's maximum principle (Theorem 6.5) for any choice x_j of orthonormal vectors

$$\sum_{j=1}^{k} \langle Bx_j, x_j \rangle \leq \sum_{j=1}^{k} \lambda_{[j]}(B).$$

The above two relations imply

$$\sum_{j=1}^{k} \lambda_{[i_j]}(A+B) \leq \min_{\substack{x_j \in M_j \\ x_j \text{ orthonormal}}} \sum_{j=1}^{k} \langle Ax_j, x_j \rangle + \sum_{j=1}^{k} \lambda_{[j]}(B).$$

Now, use Theorem 9.1 once again to conclude that the first term on the right-hand side of the above inequality is dominated by $\sum_{j=1}^{k} \lambda_{[i_j]}(A)$. ■

<u>COROLLARY 9.5</u> : For A,B Hermitian the following majorization relation between eigenvalues holds:

$$\mathrm{Eig}_{\downarrow}(A+B) - \mathrm{Eig}_{\downarrow}(A) \prec \mathrm{Eig}(B) \qquad (9.1)$$

<u>Proof</u> : Note if x and y are any two vectors such that for any $1 \leq i_1 < \ldots < i_k \leq n$ we have $x_{i_1} + \ldots + x_{i_k} \leq y_{[1]} + \ldots + y_{[k]}$ and if $x_1 + \ldots + x_n = y_1 + \ldots + y_n$ then $x \prec y$. ■

In (9.1) change B to B-A to get

$$\mathrm{Eig}_{\downarrow}(B) - \mathrm{Eig}_{\downarrow}(A) \prec \mathrm{Eig}\,(B-A). \qquad (9.2)$$

Using the HLP Theorem (Theorem 3.1) this gives an equivalent version of Lidskii's Theorem:

THEOREM 9.6 (Lidskii's theorem) : Let A,B be any two Hermitian operators on an n-dimensional space H. Then the vector $(\lambda_{[1]}(B) - \lambda_{[1]}(A),\ldots,\lambda_{[n]}(B) - \lambda_{[n]}(A))$ lies in the convex hull of the n! vectors obtained by permuting the coordinates of the vector $(\lambda_1(B-A),\ldots,\lambda_n(B-A))$.

An inequality complementary to (9.2) can be obtained by using Corollary 6.6 of the (more elementary) Ky Fan's maximum principle. It follows from there that

$$\text{Eig}\,(A+B) \prec \text{Eig}_\downarrow(A) + \text{Eig}_\downarrow(B).$$

Replace A by -A and note that $\text{Eig}_\downarrow(-A) = -\text{Eig}_\uparrow(A)$. This gives

$$\text{Eig}(B-A) \prec \text{Eig}_\downarrow(B) - \text{Eig}_\uparrow(A). \tag{9.3}$$

The majorization inequalities occurring above lead to inequalities for spectral variation in all unitarily-invariant norms. The following theorem is a very pleasing generalization of Theorem 8.5.

THEOREM 9.7 : For any two Hermitian operators A and B, we have

$$|||\text{Eig}_\downarrow(A) - \text{Eig}_\downarrow(B)||| \leq |||A-B||| \leq |||\text{Eig}_\downarrow(A) - \text{Eig}_\uparrow(B)|||$$

Proof : By (9.2) $\text{Eig}_\downarrow(A) - \text{Eig}_\downarrow(B) \prec \text{Eig}\,(A-B)$. So, by Corollary 3.4 (see the remark following it), we have

$$|\text{Eig}_\downarrow(A) - \text{Eig}_\downarrow(B)| \prec_w |\text{Eig}\,(A-B)|.$$

Since A-B is Hermitian $|\text{Eig}\,(A-B)| = \text{Sing}\,(A-B)$. So this weak majorization implies that the first inequality of the Theorem holds for every Ky Fan norm Hence by Theorem 7.5, it holds for every unitarily-invariant norm.

The second inequality is derived by applying the same argument to (9.3) instead of (9.2). ■

Using the argument employed in proving Corollary 6.7, we have

THEOREM 9.8 : Let A,B be any two nxn matrices. Then

$$|||\text{Sing}(A) - \text{Sing}(B)||| \leq |||A-B|||.$$

(Recall Sing A denotes the singular values of A arranged in decreasing order).

§10. Matrices with real eigenvalues and Lax's theorem

Matrices considered in this section will not necessarily be Hermitian, but these results fit in here naturally and so are included here. The class of matrices studied here is a real vector space each element of which has real eigenvalues. Hermitian matrices form an important subclass of this class. Matrices of this kind occur in the study of vectorial hyperbolic differential equations.

Recall that we call a matrix positive if it is Hermitian and all its eigenvalues are nonnegative. We will call a matrix A *laxly positive* if all its eigenvalues are nonnegative. (If none of these eigenvalues is zero, we will call the matrix *strictly laxly positive)*. Lax positivity of A will be symbolically denoted as $0 \leq^L A$. We will say $A \leq^L B$ (A is smaller than B in the Lax order) if B-A is laxly positive.

We will see that if $\mathcal{R}$ is a real vector space of matrices the eigenvalues of which are all real, then the laxly positive elements of $\mathcal{R}$ form a convex cone. So the Lax order $\leq^L$ defines a partial order on $\mathcal{R}$.

Given two matrices A and B we say that λ is an *eigenvalue of A with respect to B* if there exists a nonzero vector x such that $Ax = \lambda Bx$. Thus, eigenvalues of A with respect to B are the n roots of the equation

$$\det (A-\lambda B) = 0.$$

LEMMA 10.1 : Let A,B be two matrices such that every real linear combination of A and B has real eigenvalues. Suppose B is strictly laxly positive. Then for every real λ, $-A + \lambda I$ has real eigenvalues with respect to B.

Proof : We have to show that for a fixed real λ the equation

$$\det(-A + \lambda I - \mu B) = 0 \tag{10.1}$$

is satisfied by n real μ's.

Let μ be any given real number. Then by hypothesis there exist n real λ's satisfying (10.1), namely the eigenvalues of $A + \mu B$. Denote these λ's as

$$\varphi_1(\mu) \geq \varphi_2(\mu) \geq \ldots \geq \varphi_n(\mu).$$

We have,

$$\det(-A + \lambda I - \mu B) = \prod_{k=1}^{n} (\lambda - \varphi_k(\mu)). \tag{10.2}$$

As a function of μ each $\varphi_k(\mu)$ is continuous and piecewise analytic. (See notes at the end of this chapter). For large μ, $\mu^{-1}(A + \mu B)$ is close to B. Thus as μ approaches ∞, $\mu^{-1}\,\varphi_k(\mu)$ approaches the top kth eigenvalue $\lambda_{[k]}(B)$ of B; and as μ approaches $-\infty$, $\mu^{-1}\,\varphi_k(\mu)$ approaches $\lambda_{[n-k+1]}(B)$. Since B is strictly laxly positive this implies $\varphi_k(\mu) \to \pm\infty$ as $\mu \to \pm\infty$.

So for any fixed λ and for every $k = 1,2,\ldots,n$ there exists some μ such that $\lambda = \varphi_k(\mu)$. So there are n real μ satisfying (10.1). ■

PROPOSITION 10.2 : Let A,B be two matrices such that every real linear combination of A and B has real eigenvalues. Suppose A is (strictly) laxly negative. Then every eigenvalue of A + iB has (strictly) negative real part

Proof : Let $\mu = \mu_1 + i\mu_2$ be an eigenvalue of A + iB. Then $\det[A + iB - \mu_1 I - i\mu_2 I] = 0$. Multiplying by i^n this gives

$$\det[(-B + \mu_2 I) + i(A - \mu_1 I)] = 0.$$

So, the matrix $-B + \mu_2 I$ has an eigenvalue $-i$ with respect to the matrix

$A - \mu_1 I$ and it has an eigenvalue i with respect to the matrix $-(A - \mu_1 I)$.

By hypothesis, every real linear combination of $A - \mu_1 I$ and B has real eigenvalues. So by Lemma 10.1 $A - \mu_1 I$ can neither be strictly laxly positive nor strictly laxly negative. So

$$\lambda_{[n]}(A) \leq \mu_1 \leq \lambda_{[1]}(A). \tag{10.3}$$

So if $\lambda_{[1]}(A)$ is (strictly) negative then so is μ_1. ∎

Remark : We could prove in the same way that

$$\lambda_{[n]}(B) \leq \mu_2 \leq \lambda_{[1]}(B). \tag{10.4}$$

THEOREM 10.3 : Let R be a real vector space of matrices each of which has real eigenvalues. Let A,B be two elements of R such that $A \leq^L B$. Then

$$\lambda_{[k]}(A) \leq \lambda_{[k]}(B), \quad k = 1,2,\dots,n. \tag{10.5}$$

Proof : We shall prove that if $A, B \in R$ and B is laxly positive then $\lambda_{[k]}(A+B) \geq \lambda_{[k]}(A)$. A little more generally we shall show that $\lambda_{[k]}(A+\mu B)$ is monotonically increasing in the variable μ if B is laxly positive. In the notation of Lemma 10.1, $\lambda_{[k]}(A+\mu B) = \varphi_k(\mu)$. Suppose there is some μ-interval in which $\varphi_k(\mu)$ decreases. Then choose a λ such that $\lambda - \varphi_k(\mu)$ increases from a negative to a positive value in this interval. But $\varphi_k(\mu) \to \pm\infty$ as $\mu \to \pm\infty$. So, for this value of λ, $\lambda - \varphi_k(\mu)$ vanishes for at least three values of μ. So, in the representation (10.2) this factor produces at least three zeros, whereas the remaining factors contribute at least one zero each. So, for this λ, equation (10.1) has at least $n+2$ roots μ, which is impossible. ∎

THEOREM 10.4 : Let $\mathcal{R}$ be a real vector space of matrices each of which has real eigenvalues. Let $A,B \in \mathcal{R}$. Then, for $k = 1,2,\ldots,n$,

$$\lambda_{[k]}(A) + \lambda_{[n]}(B) \leq \lambda_{[k]}(A+B) \leq \lambda_{[k]}(A) + \lambda_{[1]}(B). \tag{10.6}$$

Proof : Let $c \leq \lambda_{[n]}(B)$. Then $B - cI$ is laxly positive. So, as observed in the proof of Theorem 10.3, $\lambda_{[k]}(A+\mu(B-cI)) = \lambda_{[k]}(A+\mu B) - \mu c$ is monotonicall increasing in μ. In particular, $\lambda_{[k]}(A+\mu B) - \mu\lambda_{[n]}(B)$ is monotonically increasing in μ. Choose $\mu = 0$ and 1 to get the first inequality in (10.6). The same argument shows that $\lambda_{[k]}(A+\mu B) - \mu\lambda_{[1]}(B)$ is monotonically decreasing in μ and yields the second inequality in (10.6). ■

COROLLARY 10.5 : On $\mathcal{R}$ the function $\lambda_{[1]}(A)$ is convex and the function $\lambda_{[n]}(A)$ is concave in the argument A.

THEOREM 10.6 : Let A and B be two matrices such that all their real linear combinations have real eigenvalues. Then

$$\max_k |\lambda_{[k]}(A) - \lambda_{[k]}(B)| \leq \operatorname{spr}(A-B) \leq \|A-B\|. \tag{10.7}$$

(Here spr denotes the spectral radius of a matrix).

Proof : Let $\mathcal{R}$ be the real vector space generated by A and B. Use Theorem 10.4 with $B-A$ in place of B to get

$$\lambda_{[k]}(A) + \lambda_{[n]}(B-A) \leq \lambda_{[k]}(B) \leq \lambda_{[k]}(A) + \lambda_{[1]}(B-A).$$

This gives

$$|\lambda_{[k]}(B) - \lambda_{[k]}(A)| \leq \max(|\lambda_{[1]}(B-A)|, |\lambda_{[n]}(B-A)|)$$

$$= \text{spr}(B-A) \leq \|A-B\|. \quad \blacksquare$$

Remarks : Neither of the inequalities in (10.6) holds if we are only given that A,B and A+B have real eigenvalues. For example, let

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Lax [1] gives an example to show that there exist spaces with the property assumed above which contain non Hermitian matrices.

Notes and References for Chapter 3

The basic reference for section 8 is Weyl's famous paper (Weyl [1]). Several consequences of the minmax principle have been noted by various authors. See, in particular, Chapter 10 of Parlett [1] for an excellent account which includes this material and much more.

Lidskii's Theorem has an interesting history. It seems to have been proved by Berezin and Gel'fand [1] in connection with their work on the representation theory of Lie groups. Their proof relies on the methods of representation theory. Lidskii [1] provided an elementary proof. This "elementary" proof, however, could not be clearly understood by several other mathematicians including Wielandt [2] who proved his minmax principle and then derived Lidskii's Theorem from it. There are several different proofs of Lidskii's Theorem available now. See, for example, Kato [1], Parthasarathy [1] or Bhatia and Parthasarathy [1], Simon [1].

The inequality of Theorem 9.4 has been generalized by Amir-Moez [1], [2]. The latter monograph contains several interesting results around this circle of ideas. Our proof of Wielandt's Theorem is adapted from the one given by Amir-Moez. Note that the inequality in Theorem 9.4 is not symmetric in A and B. One of the symmetric versions proved by Thompson and Freede [1] stat

$$\sum_{j=1}^{k} \lambda_{[i_j+p_j-j]}(A+B) \le \sum_{j=1}^{k} \lambda_{[i_j]}(A) + \sum_{j=1}^{k} \lambda_{[p_j]}(B)$$

for any indices with $1 \le i_1 < \ldots < i_k \le n$, $1 \le p_1 < \ldots < p_k \le n$, and $i_k + p_k - k \le n$.

Amir-Moez and also Ando [1] discuss another kind of generalization of Wielandt's theorem in which sums of eigenvalues are replaced by other functions.

For a generalization of Wielandt's minmax principle to infinite dimensions and for topological analogues of these ideas see Riddell [1].

The first part of Theorem 9.7 was explicitly recorded in Mirsky [1]. The second part was proved by Sunder [1] using a different approach from the one we have adopted here. Theorem 9.8 is proved in Mirsky [1].

Several results related to the ones presented in section 9 can be found in Marshall and Olkin [1] and in Ando [1].

The theorems in section 10 were proved by Lax [1] using the theory of linear partial differential equations of hyperbolic type. The paper of Lax was followed by one of Weinberger [1] who gave the simple matrix theoretic proofs which we have reproduced here. Gårding pointed out that these results are special cases of his results for hyperbolic polynomials (Gårding [1],[2] Two related papers are Wielandt [4] and Gerstenhaber [1].

In the proof of Lemma 10.1 we have used the fact that the eigenvalues of a matrix family $T(z) = A + zB$ can be enumerated as $\lambda_1(z),\ldots,\lambda_n(z)$ in such a way that each $\lambda_k(z)$ is continuous as a function of z and analytic except at a finite number of points. If the eigenvalues are real then the above enumeration can be taken as the descending enumeration $\lambda_1(z) \ge \ldots \ge \lambda_n(z)$ for each z. Such results can be found, e.g., in Kato [1] or in Bhatia and Parthasarathy [1].

4 Spectral variation of normal matrices

§11. Introduction to Chapter 4

In the preceding chapter, we worked with pairs of matrices each of which had real eigenvalues. This reality helped us in pairing the eigenvalues in a natural way : we arranged them in decreasing order and paired the jth eigenvalue of A with the jth eigenvalue of B. When the eigenvalues are complex, no natural order is available and this makes the spectral variation problem for arbitrary matrices more complicated.

Let Eig A = $\{\alpha_1,\ldots,\alpha_n\}$ and Eig B = $\{\beta_1,\ldots,\beta_n\}$ be the *unordered* n-tuples consisting of the eigenvalues of two nxn complex matrices A and B. As in section 1, define the ***optimal matching distance*** between Eig A and Eig B as

$$d(\text{Eig } A, \text{Eig } B) = \min_{\sigma\in S_n} \max_{1\le i\le n} |\alpha_i - \beta_{\sigma(i)}|. \tag{11.1}$$

Some other useful distances can be defined using symmetric gauge functions (unitarily-invariant norms). Let D(A) and D(B) be the diagonal matrices with $\alpha_1,\ldots,\alpha_n$ and $\beta_1,\ldots,\beta_n$ down their diagonals. Then we have

$$d(\text{Eig } A, \text{Eig } B) = \min_W \|D(A) - WD(B)W^{-1}\| \tag{11.2}$$

where, W runs over the group of permutation matrices. Write $\|(\text{Eig } A, \text{Eig } B)\|$ for d(Eig A, Eig B) and define more generally

$$|||(\text{Eig } A, \text{Eig } B)||| = \min_W |||D(A) - WD(B)W^{-1}||| \tag{11.3}$$

for any unitarily-invariant norm.

We have seen in Chapter 3 that for A, B Hermitian we have

$$|||(\text{Eig } A, \text{Eig } B)||| = |||\text{Eig}_\downarrow(A) - \text{Eig}_\downarrow(B)||| \tag{11.4}$$

and in this case

$$|||(\text{Eig } A, \text{ Eig } B)||| \leq |||A-B|||. \tag{11.5}$$

In this chapter, we study the problem of finding bounds for $|||(\text{Eig } A, \text{ Eig } B)|||$ when A and B are normal. Though our knowledge in this case is not as complete as that in the case of Hermitian matrices, several interesting results are known.

§12. The Hausdorff distance between spectra

Let $D(a,\rho)$ and $\overline{D}(a,\rho)$ denote, respectively, the open and the closed disks with centre a and radius ρ.

PROPOSITION 12.1 : Suppose A, B are normal $n\times n$ matrices with $\|A-B\| = \varepsilon$. Suppose a disk $\overline{D}(a,\rho)$ contains k eigenvalues of A. Then the disk $\overline{D}(a,\rho+\varepsilon)$ contains at least k eigenvalues of B.

Proof : Assume, without loss of generality, that $a = 0$. Suppose $\overline{D}(0,\rho)$ contains k eigenvalues of A but $\overline{D}(0, \rho+\varepsilon)$ contains less than k eigenvalues of B. Then there exists a unit vector x in the intersection of the eigenspace of A corresponding to its eigenvalues lying inside $\overline{D}(0,\rho)$ and the eigenspace of B corresponding to its eigenvalues lying outside $\overline{D}(0, \rho+\varepsilon)$. For this x, $\|Ax\| \leq \rho$, $\|Bx\| > \rho+\varepsilon$. On the other hand,

$$\|Bx\| - \|Ax\| \leq \|(B-A)x\| \leq \varepsilon.$$

This is a contradiction. ■

COROLLARY 12.2 : Let A,B be normal with $\|A-B\| = \varepsilon$. Then within a distance ε of every eigenvalue of A there is at least one eigenvalue of B and vice versa.

Proof : Choose $\rho = 0$ in the above Proposition. ■

Let Spec A denote the subset of the plane consisting of points which are eigenvalues of A. By statements 1.3 and 1.4 we have

THEOREM 12.3 : Let A,B be nxn normal matrices. Then the Hausdorff distance h(Spec A, Spec B) satisfies the inequality

$$h(\text{Spec } A, \text{Spec } B) \leq \|A-B\|.$$

When $n = 2$ we have

$$d(\text{Eig } A, \text{Eig } B) \leq \|A-B\|.$$

In the asymmetric situation when only one of the matrices is normal we have

THEOREM 12.4 : Let A be a normal and B an arbitrary nxn matrix. Let $\|A-B\| = \varepsilon$. Then every eigenvalue of B is within a distance ε of an eigenvalue of A.

Proof : Let β be any eigenvalue of B. To prove the assertion, we can assume, by applying a translation, that $\beta = 0$. Suppose every eigenvalue α_j of A is outside the disk $\overline{D}(0,\varepsilon)$. Then A is invertible and $\|A^{-1}\| = \dfrac{1}{\min |\alpha_j|} < \dfrac{1}{\varepsilon}$. So

$$\|A^{-1}(B-A)\| \leq \|A^{-1}\| \, \|B-A\| < 1.$$

Since $B = A(I+A^{-1}(B-A))$, the above inequality shows that B is invertible, but this contradicts our assumption that B has a zero eigenvalue. ■

If A,B were both normal this would give another proof of the statement of Corollary 12.2.

For later use we record

PROPOSITION 12.5 : Let A,B be nxn normal matrices. Suppose there exist two sets K_A, K_B containing, respectively, k eigenvalues of A and at least n-k+1 eigenvalues of B, and such that the convex hulls of K_A and K_B are at distanc δ from each other. Then $\delta \le \|A-B\|$.

Proof : By the hypothesis there exists a unit vector x in the intersection of the eigenspace of A for the eigenvalues in K_A and the eigenspace of B for the eigenvalues in K_B. Since A,B are normal $\langle Ax,x\rangle$ is in the convex hull of K_A and $\langle Bx,x\rangle$ is in the convex hull of K_B. So

$$\delta \le |\langle Ax,x\rangle - \langle Bx,x\rangle| \le \|A-B\|.$$

Let us also record a fact which is well-known (and which we will prove in a much stronger form in Chapter 5.

PROPOSITION 12.6 : The map $A \to \text{Eig } A$ is a continuous map from the space M(n) of matrices to the space $\mathbb{C}^n/S_n$ of unordered n-tuples of complex numbers, i.e. if $\|A_k-A\| \to 0$ then $d(\text{Eig } A_k, \text{Eig } A) \to 0$.

We remark here that the optimal matching distance is a metric on the space $\mathbb{C}^n/S_n$.

§13. Geometry and spectral variation I

We now introduce a geometric technique which will lead to several spectral variation results. Denote by $N(n)$ or N the class of all nxn normal matrices

Since tA is normal for every real t if A is normal, the set N is path connected. However, N is not an affine set.

LEMMA 13.1 : Let $A, B \in N$. Then the line segment joining A and B lies in N iff A-B is a normal matrix.

Proof : The line segment joining A and B consists of the matrices $A(t) = (1-t)A + tB$, $0 \leq t \leq 1$. One can check by an explicit computation that each A(t) is normal iff A-B is normal. ■

A continuous map $\gamma : [a,b] \to N$, where [a,b] is any interval, will be called a *normal path* or a *normal curve*. The length of γ with respect to a norm $|||\cdot|||$ is defined as

$$l_{|||\cdot|||}(\gamma) = \sup\left\{\sum_{k=0}^{m-1} |||\gamma(t_{k+1}) - \gamma(t_k)||| : a = t_0 < t_1 < \ldots < t_m = b\right\} \tag{13.1}$$

The path γ is called *rectifiable* if this length is finite. Often the path γ would be continuously differentiable (the differentiation of a matrix function is defined entrywise). In such a case

$$l_{|||\cdot|||}(\gamma) = \int_a^b |||\gamma'(t)|||dt. \tag{13.2}$$

If $\gamma(a) = A$ and $\gamma(b) = B$ we say that γ *is a path joining the matrices* A *and* B.

The key theorem of this section is

THEOREM 13.2 : Let A and B be normal matrices and let γ be a rectifiable normal path joining them. Then

$$\|(\text{Eig } A, \text{Eig } B)\| \leq 1_{\|\cdot\|}(\gamma).$$

Proof : For convenience, let us suppose that the path γ is parametrized on the interval $[0,1]$. Denote by γ_r that part of the curve which is parametrize on $[0,r]$. Let

$$G = \{r \in [0,1] : \|(\text{Eig } A, \text{Eig } \gamma(r))\| \leq 1_{\|\cdot\|}(\gamma_r)\}.$$

The theorem will be proved if we show $1 \in G$.

By the continuity of γ and of the arclength and by Proposition 12.6, the set G is closed. So, if $R = \sup G$, then $R \in G$. So, the theorem will be proved if we show that $R = 1$.

Suppose $R < 1$. Let $S = \gamma(R)$ and let d be the minimum distance between the distinct eigenvalues of the normal operator S. Using the continuity of γ and Proposition 12.6, we can find a $t \in (R,1]$ such that if $T = \gamma(t)$ then $\|(\text{Eig } S, \text{Eig } T)\| < d/2$. This inequality says that we can label the eigenvalues of S as $\lambda_1,\ldots,\lambda_n$ and those of T as $\mu_1,\ldots,\mu_n$ in such a way that $\max|\lambda_i - \mu_i| < d/2$. We claim that in this labelling each μ_i is paired with that λ_i which is closest to it, i.e. $|\lambda_i - \mu_i| \leq |\lambda_j - \mu_i|$ for all i,j. For, if this inequality is violated for some $\lambda_i \neq \lambda_j$, then for these defaulting indices $|\lambda_i - \lambda_j| \leq |\lambda_i - \mu_i| + |\mu_i - \lambda_j| < 2|\lambda_i - \mu_i| < d$, and this goes against the definition of d. Next, we claim that $\max|\lambda_i - \mu_i| \leq \|S-T\|$. For, if $|\lambda_i - \mu_i| > \|S-T\|$ for any i then $|\lambda_j - \mu_i| > \|S-T\|$ for this i and all j, which goes against Theorem 12.4. Thus $\|(\text{Eig } S, \text{Eig } T)\| \leq \|S-T\|$. But then

$$\begin{aligned}\|(\text{Eig } A, \text{Eig } \gamma(t))\| &\leq \|(\text{Eig } A, \text{Eig } S)\| + \|(\text{Eig } S, \text{Eig } T)\| \\ &\leq 1_{\|\cdot\|}(\gamma_R) + \|S-T\| \\ &\leq 1_{\|\cdot\|}(\gamma_t).\end{aligned}$$

But this would mean $t \in G$, which is not possible. ■

THEOREM 13.3 : Let A,B be normal matrices such that A-B is also normal. Then

$$\|(\text{Eig } A, \text{Eig } B)\| \leq \|A-B\|.$$

Proof : By Lemma 13.1 the path γ consisting of the line segment joining A and B is a normal path. Its length is $\|A-B\|$. ■

Note that this theorem includes Weyl's Theorem 8.1 as a special case.

To obtain spectral variation bounds from Theorem 13.2 one has to evaluate the minimal length of a normal path joining A and B. This seems a difficult problem. Theorem 13.3 concerns a very special case. In general A-B is not normal and so the straight line joining A and B goes outside N. However, it turns out that for certain A and B there exists a normal path joining them which is different from a straight line but has length $l_{\|\cdot\|}(\gamma) = \|A-B\|$, the length of the straight line joining A,B. (This is possible because the metric under consideration is not Euclidean and so geodesics are not necessarily straight lines. Note, however, that by the triangle inequality and the definition of arclength the length of any path joining A and B can not be less than $\|A-B\|$).

We will call a path $\gamma : [0,1] \to M(n)$ a *short path* if $l_{\|\cdot\|}(\gamma) = \|\gamma(0)-\gamma(1)\|$, i.e. the length of γ is the same as that of the straight line path joining the end points of γ. We will call a subset X of M(n) a *plain* if any two points in X can be joined by a short path lying within X. Our next theorem identifies one subset of normal matrices which is not an affine set but which is a plain.

THEOREM 13.4 : The set $\mathbb{C}.U(n)$ consisting of constant multiples of nxn unitary matrices is a plain.

Proof : First note that $\mathbb{C}.U(n) = \mathbb{R}_+.U(n)$. Let $N_o = r_oU_o$, $N_1 = r_1U_1$ be any two elements of this set, where r_o,r_1 are nonnegative real numbers and U_o,U_1 are unitary matrices. Choose an orthonormal basis of the underlying space in which the unitary matrix $U_1U_o^{-1}$ takes the form

$$U_1U_o^{-1} = \mathrm{diag}(e^{i\vartheta_1},\dots,e^{i\vartheta_n}), \tag{13.3}$$

where

$$|\vartheta_n| \leq \dots \leq |\vartheta_1| \leq \pi . \tag{13.4}$$

Here $\mathrm{diag}(\alpha_1,\dots,\alpha_n)$ stands for a diagonal matrix with entries $\alpha_1,\dots,\alpha_n$ down the diagonal.

By the spectral theorem, we can achieve the reduction to the form (13.3) by a unitary conjugation. Since norms as well as eigenvalues do not change under unitary conjugations we may assume that all matrices are written with respect to a basis in which $U_1U_o^{-1}$ has the above form. Let

$$K = \mathrm{diag}(i\vartheta_1,\dots,i\vartheta_n). \tag{13.5}$$

Note K is a skew-Hermitian matrix with eigenvalues lying in the interval $(-i\pi,i\pi]$. We have,

$$\begin{aligned} \|N_o - N_1\| &= \|r_oI - r_1U_1U_o^{-1}\| \\ &= \max_k |r_o - r_1 \exp(i\vartheta_k)|. \end{aligned}$$

Since r_o and r_1 are nonnegative, this gives

$$\|N_o - N_1\| = |r_o - r_1 \exp(i\vartheta_1)|, \tag{13.6}$$

which is the length of the straight line joining the points r_0 and $r_1 \exp(i\vartheta_1)$ in the plane. Parametrize this line segment as $r(t) \exp(it\,\vartheta_1)$, $0 \le t \le 1$. This can be done except when $|\vartheta_1| = \pi$, to which case we will return later. The length of this line segment is also given by

$$\int_0^1 |r'(t) \exp(it\,\vartheta_1) + r(t)\, i\vartheta_1 \exp(it\,\vartheta_1)|dt$$

$$= \int_0^1 |r'(t) + r(t)\, i\vartheta_1|dt.$$

So, we have

$$\|N_0 - N_1\| = \int_0^1 |r'(t) + r(t)\, i\vartheta_1|dt. \qquad (13.7)$$

Now define

$$N(t) = r(t).\exp(tK).U_0, \qquad 0 \le t \le 1.$$

Then $N(t)$ traces out a smooth curve in $\mathbb{C}.U(n)$ and its endpoints are N_0 and N_1. The length of this path is given by

$$l_{\|\cdot\|}(N) = \int_0^1 \|N'(t)\|dt$$

$$= \int_0^1 \|r'(t) \exp(tK)U_0 + r(t)K \exp(tK)U_0\|dt$$

$$= \int_0^1 \|r'(t)I + r(t)K\|dt. \qquad (13.8)$$

(At the last step we have used, as before, the unitary invariance of the

THEOREM 13.7 : The set $N(2)$ of 2×2 normal matrices is a plain.

Proof : Let A,B be any two elements of $N(2)$. Each has two eigenvalues. If the eigenvalues of A and the eigenvalues of B do not lie on parallel lines then an elementary geometric construction shows that they lie on concentric circles. If α is the common centre then this shows that A and B lie in $\alpha + \mathbb{C}.U(2)$ which is a plain by Theorem 13.4. If the eigenvalues lie on parallel lines then by an appropriate scalar multiplication we may assume that these lines are parallel to the real axis. In this case the skew-Hermitian part of A-B is a scalar. So A-B is normal. By Lemma 13.1 then, the line joining A and B lies in $N(2)$. ■

Note that Theorem 13.7 (together with Theorem 13.2) provide another proof of the second part of Theorem 12.3.

It is natural now to wonder whether $N(n)$ is a plain for $n > 2$. That this cannot be so is shown by the folllowing ingenious example constructed by M.D. Choi.

Example 13.8 : Consider 3×3 normal matrices

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}$$

Then $\|A-B\| = 2$. So, if there were a short normal path joining A and B then its midpoint would be a normal matrix C such that

$$\|A-C\| = \|B-C\| = 1. \tag{13.10}$$

Since each entry of a matrix is dominated in modulus by the norm of the matrix, this implies that

$$|c_{21} - 1| \leq 1 \quad \text{and} \quad |c_{21} + 1| \leq 1.$$

These two conditions force $c_{21} = 0$. The same argument shows $c_{32} = 0$. Thus

$$A-C = \begin{bmatrix} * & * & * \\ 1 & * & * \\ * & 1 & * \end{bmatrix} \qquad (13.11)$$

where the * represent entries not yet known. However, the norm of a matrix dominates the Euclidean vector norm of each row and each column. So (13.10) and (13.11) imply

$$A-C = \begin{bmatrix} 0 & 0 & * \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} .$$

So,

$$C = \begin{bmatrix} 0 & 1 & * \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} .$$

But then C cannot be normal.

Thus $N(3)$ is not a plain.

Notice that in the above example A is Hermitian and B skew-Hermitian.

§14. Geometry and spectral variation II

In this section, we will extend Theorem 13.2 to cover all unitarily-invarian norms and then reap some corollaries of the extended theorem. Whereas, the differential geometry used in section 13 was elementary, here we will need a few slightly more sophisticated notions. (Specifically, we use the notions of a manifold, the tangent space to a manifold, tangent vectors and their action on functions defined on the manifold. We use the fact that the similarity class of a matrix is a smooth manifold. This is a special case of a general theorem. All these facts are normally covered in a graduate course on differential geometry. A reader who is not familiar with these ideas could refer to one of the standard texts).

To fix the ideas, we will first work with the Frobenius norm. The space M(n) of nxn complex matrices becomes a Hilbert space with the inner product defined by

$$\langle A,B \rangle = \mathrm{tr}\, B^*A.$$

The norm arising from this inner product is the Frobenius norm $\|A\|_F = (\mathrm{tr}\, A^*\, A)^{1/2}$.

The matrix AB-BA is denoted by [A,B] and called the *Lie bracket* or the *commutator* of A and B. We denote by Z(A) the *commutant* of A in M(n), i.e.

$$Z(A) = \{X \in M(n) : [A,X] = 0\}.$$

Note this is a subspace of M(n).

Let GL(n) be the multiplicative group of all nxn invertible matrices. The *adjoint action* of GL(n) on M(n) is the map $A \to gAg^{-1}$ defined for $A \in M(n)$, $g \in GL(n)$. The *orbit* of A under this action is the set

$$O_A = \{gAg^{-1} : g \in GL(n)\}$$

consisting of all matrices similar to A. This set is a smooth submanifold in the Hilbert space M(n). The *tangent space* to O_A at the point A will be denoted by $T_A\, O_A$. This is a linear subspace of M(n). The following proposition identifies $T_A\, O_A$ and its orthogonal complement in the Hilbert space M(n).

PROPOSITION 14.1 : Let $A \in M(n)$. Then

$$T_A\, O_A = \mathrm{span}\, \{[A,X] : X \in M(n)\}$$

$$(T_A\, O_A)^{\perp} = Z(A^*),$$

where $\perp$ denotes the orthogonal complement of a subspace in the Hilbert space M(n).

Proof : A differentiable curve in O_A passing through A can be parametrized as

$$\lambda(t) = g(t)\, Ag(t)^{-1}, \quad \lambda(0) = A.$$

Note that these equations imply that $A = g(0)^{-1}\, Ag(0)$ and hence,

$$\lambda(t) = h(t)\, Ah(t)^{-1}, \text{ where } h(t) = g(t)\, g(0)^{-1}.$$

Thus every differentiable curve in O_A passing through A has the form

$$\lambda(t) = h(t)\, Ah(t)^{-1}, \quad \text{where} \quad h(0) = I.$$

Differentiating at $t = 0$ gives

$$\dot{\lambda}(0) = \dot{h}(0)A - A\dot{h}(0) = [\dot{h}(0),A].$$

Thus tangent vectors to O_A at the point A can be written as commutators $[A,X]$. Further, every such commutator can be obtained as a tangent vector to the curve $\lambda(t) = \exp(tX)A\exp(-tX)$. The tangent space $T_A O_A$, by definition, is the linear span of all these tangent vectors.

Now note that B is orthogonal to this space iff for all $X \in M(n)$ we have

$$0 = \langle [A,X],B\rangle = \operatorname{tr} B^*(AX-XA)$$

$$= \operatorname{tr}(B^*A - AB^*)X = \langle [B^*,A],X^*\rangle.$$

This is possible iff $[B^*,A] = 0$, i.e., iff $B \in Z(A^*)$. ■

For each A, we thus have a direct sum decomposition

$$M(n) = T_A O_A \oplus Z(A^*). \tag{14.1}$$

Recall that A is normal iff $Z(A) = Z(A^*)$. So,

$$M(n) = T_A O_A \oplus Z(A) \quad \text{if} \quad A \text{ is normal.} \tag{14.2}$$

Let $B \in Z(A)$. Then there exists a unitary U such that $UAU^{-1} = T(A)$, $UBU^{-1} = T(B)$, where $T(A)$, $T(B)$ are upper triangular matrices. Since the diagonal entries of these triangular matrices are the eigenvalues of A and B it follows that

$$\|(\text{Eig } A, \text{Eig } B)\|_F \leq \|T(A) - T(B)\|_F = \|A-B\|_F, \tag{14.3}$$

if B is any matrix in Z(A).

Also note that if $B \in O_A$ then Eig A = Eig B. So,

$$\|(\text{Eig } A, \text{Eig } B)\|_F = 0 \quad \text{if} \quad B \in O_A. \tag{14.4}$$

Relations (14.2), (14.3) and (14.4) suggest that we can try to estimate the spectral variation of a normal matrix componentwise along two complementary directions. A little more precisely, let A_0 be a given normal matrix and let $\gamma(t)$, $0 \leq t \leq 1$ be a normal curve with $\gamma(0) = A_0$. Consider the function $\varphi(A) = \|(\text{Eig } A_0, \text{Eig } A)\|_F$ defined for every matrix A. At each point $\gamma(t)$ consider the decomposition $M(n) = T_{\gamma(t)} O_{\gamma(t)} \oplus Z(\gamma(t))$. As we move along the curve γ, the rate of change of φ is zero in the first direction because of (14.4), in the second direction it does not change faster than the argument because of (14.3). So it is "obvious" that for $0 \leq t \leq 1$

$$\varphi(\gamma(t)) \leq \int_0^t \|\gamma'(s)\|_F \, ds.$$

In particular, if $\gamma(1) = A_1$ then

$$\|(\text{Eig } A_0, \text{Eig } A_1)\|_F \leq \text{the Frobenius length of the path } \gamma.$$

We will now prove this statement and extend it to all unitarily-invariant norms.

First of all note that the decomposition (14.2) of M(n) is valid, as a vector space direct sum decomposition, irrespective of the norm used. Inequality (14.3) extends to all unitarily-invariant norms because of Theorem 7.8,and (14.4) is true for all norms trivially.

LEMMA 14.2 : Let φ be a real valued C^1 function on a Banach space X. Let $\gamma : [0,1] \to X$ be a piecewise C^1 curve. Suppose

(i) $\gamma(0) = x_0$, $\gamma(1) = x_1$, $\varphi(x_0) = 0$;

(ii) for every t in [0,1] the space X (which is also the tangent space $T_{\gamma(t)} X$ in our notation) splits as $X = T^{(1)}_{\gamma(t)} \oplus T^{(2)}_{\gamma(t)}$ in such a way that the directional derivatives $v^{(1)}\varphi$ and $v^{(2)}\varphi$ of φ in these two directions satisfy the conditions

$$v^{(1)}\varphi = 0 \qquad \text{for all} \quad v^{(1)} \in T^{(1)}_{\gamma(t)}$$

$$v^{(2)}\varphi \leq \|v^{(2)}\| \quad \text{for all} \quad v^{(2)} \in T^{(2)}_{\gamma(t)}.$$

Let $P^{(1)}_t$, $P^{(2)}_t$ denote the complementary projections in X onto the subspaces $T^{(1)}_{\gamma(t)}$ and $T^{(2)}_{\gamma(t)}$ respectively. Then,

$$\varphi(x_1) \leq \int_0^1 \|P^{(2)}_t \; \gamma'(t)\| dt.$$

Proof : We have

$$\varphi(x_1) = \int_0^1 \gamma'(t)(\varphi) dt$$

$$= \int_0^1 (P^{(1)}_t \; \gamma'(t))(\varphi) dt + \int_0^1 (P^{(2)}_t \; \gamma'(t))(\varphi) dt$$

$$\leq 0 + \int_0^1 \|P_t^{(2)} \gamma'(t)\| dt. \quad \blacksquare$$

The statement of this lemma remains valid if φ is C^1 on a dense open set G in X and γ is a piecewise C^1 curve which intersects the complement of G at a finite number of points. In such a case we say that φ is *generically* C^1 and γ is a curve *adapted* to φ.

Let X be the space M(n) with any of the norms $|||\cdot|||$. Let A_0 be a fixed matrix and let $\varphi(A) = |||(\text{Eig } A_0, \text{Eig } A)|||$. Then φ is a generically C^1 function. (It will fail to be differentiable at those points where the differences of the eigenvalues of A_0 and A do not have distinct modulii or where the minimum in the definition of $|||(\text{Eig } A_0, \text{Eig } A)|||$ is attained at two different permutations. Such matrices A form a closed nowhere dense set in M(n)).

We can now prove the key theorem of this section:

<u>THEOREM 14.3</u>: Let M(n) be the space of matrices with any of the unitarily-invariant norms $|||\cdot|||$. Let $A : [0,1] \to M(n)$ be a piecewise C^1 curve with the following properties

(i) A(t) is normal for all $0 \leq t \leq 1$,

(ii) $A(0) = A_0$, $A(1) = A_1$,

(iii) A(t) is adapted to the generically C^1 function $\varphi(A) = |||(\text{Eig } A_0, \text{Eig } A)|||$.

Let $P_t^{(1)}$ and $P_t^{(2)}$ denote the complementary projection operators in M(n) corresponding to the direct sum decomposition $M(n) = T_{A(t)} O_{A(t)} \oplus Z(A(t))$. Then

$$\|(\text{Eig } A_0, \text{Eig } A_1)\| \leq \int_0^1 |||P_t^{(2)} A'(t)||| dt \leq \int_0^1 |||A'(t)||| dt, \qquad (14.5)$$

where A'(t) denotes the derivative of A(t).

Proof : We apply Lemma 14.2 to the Banach space M(n), the function $\varphi(A)$ and the curve A(t). Let $T^{(1)}_{A(t)} = T_{A(t)}\, O_{A(t)}$, $T^{(2)}_{A(t)} = Z(A(t))$. Choose and fix a point s in [0,1]. For every B in $O_{A(s)}$ we have $\varphi(B) = \varphi(A(s))$. So, the derivative of φ in the direction of $O_{A(s)}$ is zero, i.e.,

$$v^{(1)}\varphi = 0 \quad \text{for all} \quad v^{(1)} \in T^{(1)}_{A(s)} .$$

Define $\psi(A) = |||(\text{Eig } A(s), \text{Eig } A)|||$, and put
$h(A) = \varphi(A(s)) + \psi(A) = |||(\text{Eig } A_0, \text{Eig } A(s))||| + |||(\text{Eig } A(s), \text{Eig } A)|||$.
Note that $\varphi(A(s)) = h(A(s))$ and $\varphi(A) \leq h(A)$ for all $A \in M(n)$. Hence

$$v^{(2)}\varphi \leq v^{(2)}h \quad \text{for all} \quad v^{(2)} \in T^{(2)}_{A(s)} .$$

Also, note

$$v^{(2)}h = v^{(2)}\psi \quad \text{for all} \quad v^{(2)} \in T^{(2)}_{A(s)}$$

and

$$v^{(2)}\psi \leq |||v^{(2)}||| \quad \text{for all} \quad v^{(2)} \in T^{(2)}_{A(s)}.$$

The last statement follows from the inequality (14.3) (which, we noted, holds for any $|||\cdot|||$) and the fact that $T^{(2)}_{A(s)} = Z(A(s))$. so, we have

$$v^{(2)}\varphi \leq |||v^{(2)}||| \quad \text{for all} \quad v^{(2)} \in T^{(2)}_{A(s)}.$$

So, from Lemma 14.2 we have

$$\varphi(A_1) \leq \int_0^1 |||P_t^{(2)} A'(t)|||dt.$$

This proves the first inequality in (14.5). To prove the second we claim that $|||P_t^{(2)} B||| \leq |||B|||$ for all B. Choose a basis in which the normal matrix A(t) is diagonal. Then Z(A(t)) consists of block-diagonal matrices and $P_t^{(2)}$ B is the pinching of B by the spectral projections of A(t). So our claim follows from Theorem 7.8. ■

Remark 14.4 : To sum up, we have proved that (subject to some technical restrictions) if A and B are two normal matrices and γ a piecewise C^1 curve joining them and passing only through normal matrices, then

$$|||(\text{Eig } A, \text{Eig } B)||| \leq l_{|||\cdot|||}(\gamma). \tag{14.6}$$

This generalizes Theorem 13.2.

Now note that by continuity the above inequality would hold for all matrices if it holds on an everywhere dense set of matrices. So by perturbing our given matrices slightly we may assume that the technical restrictions stipulated in Theorem 14.3 are satisfied by A and B. Since our concern is with inequalities of type 14.6 we will, henceforth, assume these conditions are always satisfied.

The plodding is over. We can now enjoy the fruits of our labour.

THEOREM 14.5 : Let A,B be normal matrices such that A-B is also normal. Then for every unitarily-invariant norm we have

$$|||(\text{Eig } A, \text{Eig } B)||| \leq |||A-B|||.$$

Proof : The same as that of Theorem 13.3. ■

Note that this Theorem applies, as a special case, to Hermitian A and B.

THEOREM 14.6 : Let A,B be unitary matrices and let K be any skew-Hermitian matrix such that $BA^{-1} = \exp K$. Then for every unitarily-invariant norm:

$$|||(\mathrm{Eig}\ A,\ \mathrm{Eig}\ B)||| \leq |||K|||.$$

Proof : Let $A(t) = (\exp tK)A$, $0 \leq t \leq 1$. Then $A(t)$ is unitary, $A(0) = A$, $A(1) = B$, $A'(t) = K(\exp tK)A$, $|||A'(t)||| = |||K|||$. So, the length of the path $A(t)$ is

$$\int_0^1 |||A'(t)|||dt = |||K|||. \quad ■$$

THEOREM 14.7 : Let A,B be unitary matrices. Then for every unitarily-invariant norm

$$|||(\mathrm{Eig}\ A,\ \mathrm{Eig}\ B)||| \leq \frac{\pi}{2}\ |||A-B||| \tag{14.7}$$

Proof : In view of Theorem 14.6 we need to show

$$\inf\{|||K||| \ :\ BA^{-1} = \exp K\} \leq \frac{\pi}{2}\ |||A-B|||.$$

Choose a K whose eigenvalues are contained in the interval $(-i\pi, i\pi]$. Assume by applying a unitary conjugation that

$$K = \mathrm{diag}(i\vartheta_1, \ldots, i\vartheta_n).$$

Then

$$|||A-B||| = |||I-BA^{-1}||| = |||\mathrm{diag}(1-e^{i\vartheta_1},\ldots,1-e^{i\vartheta_n})|||.$$

Now note that $|\vartheta| \leq \frac{\pi}{2}|1-e^{i\vartheta}|$ for all $\vartheta \in (-\pi,\pi]$. Recall that every unitarily-invariant norm is a symmetric gauge function of singular values and every symmetric gauge function is monotonically increasing (Corollary 7.2). It follows that $|||K||| \leq \frac{\pi}{2}|||A-B|||$. ■

We now give an example to show that the constant in the inequality (14.7) cannot be replaced by a smaller constant if the inequality is to hold for all unitarily-invariant norms. (We saw in section 13 that if the operator norm alone is involved we can replace $\pi/2$ by 1).

Example 14.8 : Choose the trace norm

$$\|A\|_{tr} = \sum_{j=1}^{n} s_j(A).$$

Let A_+ and A_- be the unitary matrices obtained by adding an entry ±1 in the bottom left corner to an upper Jordan nilpotent matrix, i.e.,

$$A_\pm = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & 1 \\ \pm 1 & 0 & 0 & \ldots & 0 \end{bmatrix}$$

Then $\|A_+ - A_-\|_{tr} = 2$. The eigenvalues of $A_\pm$ are the n roots of ±1. One can see that $\|(\mathrm{Eig}\, A_+, \mathrm{Eig}\, A_-)\|_{tr}$ approaches π as $n \to \infty$.

Finally, we digress from our main concern and give a proof of the following famous theorem:

THEOREM 14.9 : Every complex matrix with trace zero can be expressed as a commutator of two matrices.

Proof : We have to show that if tr A = 0 then there exist two matrices B and C such that $A = [B,C]$. Since we can find a unitary U such that UAU^{-1} is upper triangular and since $U[B,C]U^{-1} = [UBU^{-1}, UCU^{-1}]$, it is enough to prove this for an upper triangular matrix A.

We will show that if A is upper triangular with tr A = 0 then $A = [B,C]$ where B is the nilpotent upper Jordan matrix

$$B = \begin{bmatrix} 0 & 1 & 0 \dots 0 \\ 0 & 0 & 1 \dots 0 \\ \cdot & \cdot & \cdot \quad \cdot \quad \cdot \\ 0 & 0 & 0 \dots 1 \\ 0 & 0 & 0 \dots 0 \end{bmatrix}.$$

By Proposition 14.1 we need to show that A belongs to the orthogonal complement of $Z(B^*)$. Now $Z(B^*)$ contains only polynomials in B^*. (This is a general fact : the commutant of X consists of polynomials in X iff the matrix X is non derogatory, i.e. in its Jordan decomposition there is just one block for each distinct eigenvalue). Thus $Z(B^*)$ consists of matrices of the form

$$X = \begin{bmatrix} \alpha_1 & 0 & & \dots 0 \\ \alpha_2 & \alpha_1 & & \dots 0 \\ \alpha_3 & \alpha_2 & \alpha_1 & \dots 0 \\ \alpha_n & \dots & \alpha_2 & \alpha_1 \end{bmatrix}$$

i.e. lower triangular matrices with the same entry on each sub-diagonal. Now if $A = ((a_{ij}))$ is upper triangular with trace zero, then for X as above

$$\text{tr } A^*X = 0.$$

So $A \in Z(B^*)^{\perp}$. ■

§15. The Hoffman-Wielandt theorem

The spectral variation problem for arbitrary normal matrices in the Frobenius norm is completely solved. We have:

THEOREM 15.1 (Hoffman and Wielandt) : Let A,B be normal matrices. Then

$$\|(\mathrm{Eig}\ A,\ \mathrm{Eig}\ B)\|_F \le \|A-B\|_F .$$

Proof : Let $\alpha_1,\ldots,\alpha_n$ and $\beta_1,\ldots,\beta_n$ be the respective eigenvalues of A and B. We have to show

$$\min_{\sigma} \left(\sum_{i=1}^{n} |\alpha_i - \beta_{\sigma(i)}|^2 \right)^{1/2} \le \|A-B\|_F \tag{15.1}$$

where the minimum is taken over all permutations. Write $A = \sum_{i=1}^{n} \alpha_i P_i$, $B = \sum_{i=1}^{n} \beta_i Q_i$, where P_i, Q_i are the one-dimensional eigenprojections of A,B respectively. Then

$$\|A-B\|_F^2 = \sum_{i=1}^{n} |\alpha_i|^2 + \sum_{i=1}^{n} |\beta_i|^2 - 2\ \mathrm{Re} \sum_{i,j} \alpha_i \overline{\beta}_j \ \mathrm{tr}\ P_i Q_j . \tag{15.2}$$

Put $d_{ij} = \mathrm{tr}\ P_i Q_j$, $i,j = 1,2,\ldots,n$. Then the matrix $D = ((d_{ij}))$ is doubly stochastic. So by Birkhoff's Theorem (Theorem 2.1) we can write D as a convex combination of permutation matrices, i.e.

$$D = \sum_{\sigma \in S_n} a_\sigma \sigma \quad , \qquad \sum_{\sigma \in S_n} a_\sigma = 1,$$

where $a_\sigma \ge 0$. We will use the same symbol σ to denote a permutation on n symbols and also the corresponding permutation matrix.

Let $\alpha = (\alpha_1,\ldots,\alpha_n)$ be the vector with n coordinates α_i and $\|\alpha\| = (|\alpha_1|^2 + \ldots + |\alpha_n|^2)^{1/2}$ its Euclidean vector norm. Then (15.2) can b written as

$$\|A-B\|_F^2 = \sum_\sigma a_\sigma\{\|\alpha\|^2 + \|\beta\|^2 - 2 \operatorname{Re} \langle\alpha,\sigma(\beta)\rangle\}$$

$$= \sum_\sigma a_\sigma\{\|\alpha - \sigma(\beta)\|^2\}$$

$$\geq \min_\sigma \|\alpha - \sigma(\beta)\|^2$$

$$= \min_\sigma \sum_i |\alpha_i - \beta_{\sigma(i)}|^2. \quad \blacksquare$$

<u>REMARK 15.2</u> : The above argument also gives an equality complementary to (15.1) :-

$$\|A-B\|_F \leq \max_\sigma (\Sigma|\alpha_i - \beta_{\sigma(i)}|^2)^{1/2}. \qquad (15.3$$

<u>REMARK 15.3</u> : In general there is no good prescription for describing the permutation σ for which the minimum in (15.1) is attained. However, if A is Hermitian and B arbitrary normal then the minimum is attained for the ordering in which

$$\alpha_1 \geq \ldots \geq \alpha_n; \quad \operatorname{Re} \beta_1 \geq \ldots \geq \operatorname{Re} \beta_n.$$

To see this one only has to note that if $\alpha_1 \geq \alpha_2$ and $\operatorname{Re} \beta_1 \geq \operatorname{Re} \beta_2$, then

$$|\alpha_1 - \beta_1|^2 + |\alpha_2 - \beta_2|^2 \leq |\alpha_1 - \beta_2|^2 + |\alpha_2 - \beta_1|^2 .$$

REMARK 15.4 : The conclusion of Theorem 15.1 is no longer valid if only A is normal (even Hermitian) but B is arbitrary. For example, if

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},$$

then, $\|(\text{Eig } A, \text{Eig } B)\|_F = \sqrt{2}$, $\|A-B\|_F = 1$.

§16. Perturbation of spectral spaces

In these notes we have not touched upon the problem of perturbation of eigenvectors or eigenspaces. In this section we will use one result in that direction to derive an eigenvalue variation bound.

As seen in section 14, the inequality of Theorem 15.1 does not extend to all unitarily-invariant norms even when A,B are unitary. However, in section 13 we saw that for the case of unitary matrices, the corresponding inequality holds for the operator norm. One would like to extend this analogue of Weyl's Inequality to all normal operators. Notice that since $\|A\| \leq \|A\|_F \leq \sqrt{n}\, \|A\|$, the Hoffman-Wielandt theorem gives

$$\|(\text{Eig } A, \text{Eig } B)\| \leq \sqrt{n}\, \|A-B\|,$$

for normal A,B. (Actually n here can easily be replaced by the rank of A-B). In this section we will show that in this inequality $\sqrt{n}$ can be replaced by a universal constant independent of the dimension and of A,B. To come to this result we will need some results on operator equations and spectral subspaces which are of independent interest.

THEOREM 16.1 : Let A,B be any two nxn matrices with disjoint spectra (sets of eigenvalues). Then given any matrix S the equation

$$AQ - QB = S \qquad (16.1)$$

has a unique solution Q.

Proof : Consider the operators L_A and R_B defined on $M(n)$ by $L_A(Q) = AQ$, $R_B(Q) = QB$. If x is an eigenvector of A for the eigenvalue α then the matrix X with one column x and the rest of the columns zero is an eigenvector of L_A for the eigenvalue α. Thus the eigenvalues of the operator L_A are the eigenvalues of A each with multiplicity n times as much. This accounts for all the n^2 eigenvalues of L_A. In the same way the eigenvalues of R_B are the eigenvalues of B each with multiplicity n times as much. Let $I_{A,B} = L_A - R_B$. Since L_A and R_B commute

$$\begin{aligned}\text{Spec } (I_{A,B}) &\subset \text{Spec } (L_A) - \text{Spec } (R_B)\\ &= \text{Spec } (A) - \text{Spec } (B).\end{aligned}$$

So if the spectra of A and B are disjoint, then zero is not an eigenvalue of $I_{A,B}$. So $I_{A,B}$ is an invertible operator. This proves the theorem. ■

We will now write the solution Q of (16.1) when A,B are normal matrices. Let

$$\delta = \text{dist (Spec } A, \text{ Spec } B). \tag{16.2}$$

Write $A = A_1 + iA_2$ where A_1, A_2 are commuting Hermitian matrices. In the same way write $B = B_1 + iB_2$.

We will need to use Fourier transforms in $\mathbb{R}^2$. As usual we write points in the plane as $s = (s_1, s_2)$, $t = (t_1, t_2)$, etc. The inner product $s.t$ is defined by $s.t = s_1t_1 + s_2t_2$. If f is any function in $L_1(\mathbb{R}^2)$ its Fourier transform is defined by

$$\hat{f}(t) = \iint_{\mathbb{R}^2} \exp(-it.x)\, f(x)dx.$$

Define two-parameter unitary groups $U(t)$, $V(t)$ by

$$U(t) = \exp[i(t_1A_1 + t_2A_2)], \quad V(t) = \exp[i(t_1B_1 + t_2B_2)] \tag{16.3}$$

(Note that since A_1, A_2 are commuting Hermitian matrices, U(t) is unitary for every t and the map $t \to U(t)$ is a group homomorphism from the additive group $\mathbb{R}^2$ to the multiplicative group of unitary matrices).

THEOREM 16.2 : If A and B are normal matrices with disjoint spectra then the solution to the equation (16.1) is given by

$$Q = \iint_{\mathbb{R}^2} U(-t)\ SV(t)\ f_\delta(t)dt \tag{16.4}$$

where U(t), V(t) are defined by (16.3), $\delta > 0$ is the number defined by (16.2) and f_δ is any function in $L_1(\mathbb{R}^2)$ whose Fourier transform has the property

$$\hat{f}_\delta(t) = \frac{1}{t_1 + it_2} \quad \text{for} \quad |t| \geq \delta. \tag{16.5}$$

Proof : Let u be an eigenvector of A for the eigenvalue α,v an eigenvector of B for the eigenvalue β. Write $\alpha = \alpha_1 + i\alpha_2$, $\beta = \beta_1 + i\beta_2$. (This is in conflict with our earlier notation but should not cause any confusion since this notation will be dropped after this proof). Note $A_1u = \alpha_1u$, $A_2u = \alpha_2u$, $U(t)u = e^{it.\alpha}u$, etc.

Since eigenvectors of A and B form orthonormal bases for the underlying Hilbert space it will be enough to prove that if Q is defined by (16.4) then for any such u,v

$$\langle (AQ - QB)v,u\rangle = \langle Sv,u\rangle. \tag{16.6}$$

Note that

$$\langle AU(-t)\ SV(t)v,u\rangle = \langle SV(t)v,U(t)\ A^*u\rangle$$

$$= \langle Se^{i\beta.t}v,\ e^{i\alpha.t}\ (\alpha_1 - i\alpha_2)u\ \rangle$$

$$= (\alpha_1 + i\alpha_2)\ e^{i(\beta-\alpha).t} \langle Sv,u\rangle.$$

In the same way

$$\langle U(-t)\ SV(t)Bv,u\rangle = (\beta_1 + i\beta_2)\ e^{i(\beta-\alpha).t} \langle Sv,u\rangle\ .$$

So if Q is defined by (16.4) then

$$\langle (AQ-QB)v,u\rangle = \langle Sv,u\rangle\ (\alpha_1 - \beta_1 + i(\alpha_2 - \beta_2)) \iint_{\mathbb{R}^2} e^{i(\beta-\alpha).t}\ f_\delta(t)dt$$

$$= \langle Sv,u\rangle$$

by (16.5) and the fact that $|\beta-\alpha| \geq \delta$. ∎

(We assumed above the existence of a function f_δ in $L_1(\mathbb{R}^2)$ whose Fourier transform satisfies the condition (16.5). Such functions indeed exist. We will briefly indicate how to prove this at the end of the chapter).

<u>THEOREM 16.3</u> : Let A,B be normal operators with dist(Spec A, Spec B) = $\delta > 0$ Let Q be the unique solution of the equation (16.1). Then

$$\|Q\| \leq \frac{c_2}{\delta}\ \|S\| \tag{16.7}$$

where c_2 is the constant defined as

$$c_2 = \inf\{\|f\|_{L_1} : f \in L_1(\mathbb{R}^2)\ ,\ \hat{f}(t) = \frac{1}{t_1 + it_2} \text{ if } |t| \geq 1\}. \tag{16.8}$$

Proof : Use the representation (16.4) and the unitary-invariance of the norm to conclude

$$\|Q\| \leq \|S\| \iint_{\mathbb{R}^2} |f_\delta(t)|dt$$

$$= \|S\| \; \|f_\delta\|_{L_1} \; .$$

A change of variables in f leads to the statement of the theorem. ∎

THEOREM 16.4 : Let A,B be two normal operators and let K_A, K_B be two subsets of the plane such that $\text{dist}(K_A,K_B) = \delta > 0$. Let E be the spectral projection of A corresponding to its eigenvalues lying in K_A and F the spectral projection of B corresponding to its eigenvalues lying in K_B. Let Q be any operator. Then

$$\|EQF\| \leq \frac{c_2}{\delta} \|E(QA - QB)F\| \tag{16.9}$$

where c_2 is the constant defined in (16.8).

Proof : Let $E^\perp$ and $F^\perp$ denote projections orthogonal to E and F respectively. Then for a sufficiently large constant α, we have

$$\text{dist}(\text{Spec}\,(EA + \alpha E^\perp), \text{Spec}(BF - \alpha F^\perp)) = \delta.$$

Note that the projections E and F commute with A and B respectively. So

$$(EA + \alpha E^\perp).EQF - (EQF).(BF - \alpha F^\perp) = E(AQ - QB)F.$$

The result now follows from Theorem 16.3. ∎

REMARK 16.5 : The special case when $Q = I$ is the one we will use. We have under the above conditions:

$$\|EF\| \leq \frac{c_2}{\delta} \|E(A-B)F\| \leq \frac{c_2}{\delta} \|A-B\|. \tag{16.10}$$

This inequality gives us a bound for the "angle" between two spectral subspaces E and F belonging to some parts of the spectra of A and B, in terms of $\|A-B\|$ and the separation δ of the relevant parts of the spectra.

As a corollary we have the following spectral variation result:

THEOREM 16.6 : There exists a universal constant c such that for any two normal matrices A and B

$$\|(\text{Eig } A, \text{ Eig } B)\| \leq c\|A-B\|.$$

Further $c \leq c_2$, where c_2 is defined by (16.8).

Proof : Let A,B be given normal matrices with eigenvalues $\alpha_1,\ldots,\alpha_n$ and $\beta_1,\ldots,\beta_n$ respectively. Put $\eta = c_2\|A-B\|$. We have to show that there is a permutation σ such that

$$\max_i |\alpha_i - \beta_{\sigma(i)}| \leq \eta \tag{16.11}$$

Suppose this is not the case. Then there exists $\delta > \eta$ such that for every permutation σ there exists an index i such that $|\alpha_i - \beta_{\sigma(i)}| > \delta$.

We now appeal to the Marriage Theorem (see section 1). Let $B = \{\alpha_1,\ldots,\alpha_n\}$, $G = \{\beta_1,\ldots,\beta_n\}$ and define a relation $R \subset B \times G$ by stipulating that $(\alpha_i,\beta_j) \in R$ iff $|\alpha_i - \beta_j| \leq \delta$. Then the statement of the above paragraph implies that the society (B,G,R) is not espousable. So there exist some k indices $i_1,\ldots,i_k$ such that $|G_{i_1\ldots i_k}| < k$. In other words if $K_A = \{\alpha_{i_1},\ldots,\alpha_{i_k}\}$ then the set of those β which are within a distance δ from some element of K_A has cardinality less than k. Let K_B be those β which are not in $G_{i_1 \ldots i_k}$. Then K_B contains at least $n - k + 1$ eigenvalues of B, and dist $(K_A, K_B) \geq \delta$.

Let E, F be the spectral projections of A, B corresponding to the sets K_A, K_B respectively. Then by (16.10) we have

$$\|EF\| \leq \frac{c_2}{\delta} \|A-B\| = \frac{\eta}{\delta} < 1.$$

But E is a projection of rank k and F is a projection of rank at least $n - k + 1$. So there is a common nonzero vector in their range spaces. Hence $\|EF\| = 1$. This is a contradiction. ■

REMARK 16.7 : It is possible (and likely) that the constant c of Theorem 16.6 is strictly smaller than c_2. The exact value of c_2 is not known. There are some unpublished results of Koosis which show that $c_2 < \pi$.

§17. The cyclic order

So far in this chapter we have not attempted to prescribe any order on the eigenvalues of A and B for which the optimal matching between them could be attained. This is in contrast with the results in Chapter 3 where the eigenvalues, being real, could be naturally ordered in a descending order. In one interesting case - the unitary matrices - the eigenvalues lie on the unit circle, where the cyclic order presents itself naturally. We will now briefly indicate how this order can be exploited to give another proof of Theorem 13.6. This proof is not only completely different but also reveals more.

Let γ_1, γ_2 be two points on the unit circle. We will write $\gamma_1 < \gamma_2$ if the minor arc from γ_1 to γ_2 goes counter-clockwise.

The following special result is easier to prove than the general result and is illuminating:

THEOREM 17.1 : Let A, B be two unitary matrices with all their eigenvalues lying in one semi-circle. Arrange the eigenvalues of A and B as $\alpha_1 \leq \ldots \leq \alpha_n$ and $\beta_1 \leq \ldots \leq \beta_n$ respectively. Then

$$\max_{1 \leq i \leq n} |\alpha_i - \beta_i| \leq \|A-B\|.$$

Proof : Suppose the maximum on the left-hand side is attained at $i = j$ and is equal to δ. Let $K_A = \{\alpha_1, \ldots, \alpha_j\}$, $K_B = \{\beta_j, \ldots, \beta_n\}$. Assume, without loss of generality, $\alpha_j < \beta_j$. The sets K_A, K_B in the plane can be separated by two parallel straight lines at distance δ. Apply Proposition 12.5 now. ∎

For the case of arbitrary unitary matrices this argument no longer works, but it can be modified. We cannot explicitly prescribe an order now but it turns out that one permutation for which the optimal matching distance is attained arranges the α's and the β's both in a cyclic order.

Another proof of Theorem 13.6 : The case $n = 2$ is trivial. We assume $n \geq 3$. Assume for simplicity that all the α_i and β_i are distinct and that all the distances $|\alpha_i - \beta_j|$ are distinct. Matrices with such eigenvalues are dense, so the general case would follow from this by continuity.

We will write $(\alpha\beta\gamma)$ to mean that the points α, β, γ are in counter-clockwise cyclic order on the unit circle. We will adopt a similar notation for more than three points. Indices will be numbered modulo n, thus $\alpha_{n+1} = \alpha_1$, etc. Number the α_i so that $(\alpha_1\alpha_2 \ldots \alpha_n)$. Let

$$\delta = \min_{\sigma} \max_{i} |\alpha_i - \beta_{\sigma(i)}|.$$

Assume $\delta < 2$, otherwise we have nothing to prove. Let the eigenvalues of B be numbered $\beta_1, \ldots, \beta_n$ in such a way that for any subset J of $\{1,2,\ldots,n\}$ and for any permutation σ of J we have

$$\max_{i \in J} |\alpha_i - \beta_i| \leq \max_{i \in J} |\alpha_i - \beta_{\sigma(i)}|. \quad (17.1)$$

This implies that $\max_{1 \leq i \leq n} |\alpha_i - \beta_i| = \delta$. Assume, without loss of generality that this maximum is attained for i = 1 and further that $\alpha_1 < \beta_1$ (in the notation introduced earlier in this section).

Now the following facts can be verified:

(i) If for any i, $\beta_i < \alpha_i$ then neither $(\alpha_1\ \beta_i\ \beta_1)$ nor $(\alpha_1\ \alpha_i\ \beta_1)$. To see this apply (17.1) to indices 1 and i.

(ii) There exists j such that $|\alpha_{j+1} - \beta_j| > \delta$. If not, then we could have paired each α_{i+1} with β_i to get an optimal matching distance smaller than δ.

Choose and fix one such j.

(iii) Apply (i) and (17.1) to see that $(\alpha_1\ \beta_1\ \beta_j\ \alpha_{j+1})$.

(iv) For $1 < i < j$ we have $(\beta_1\ \beta_i\ \beta_j)$. This can be seen separately for the several different cases which can arise. Verify this first for the case $\alpha_j < \beta_j$. This is easy. Let $\beta_j < \alpha_j$. First note that in this case (iii) and the condition $|\beta_j - \alpha_j| < \delta$ imply the configuration $(\alpha_1\ \beta_1\ \beta_j\ \alpha_j\ \alpha_{j+1})$. Now consider the two subcases $\alpha_i < \beta_i$ and $\beta_i < \alpha_i$ separately. In each case (iv) has to be true.

Now let K_A be the arc from α_{j+1} positively to α_1 and K_B the arc from β_1 positively to β_j. By the original numbering of the α_i there are n - j + 1 of the α_i lying in K_A. By (iv) above there are j of the β_i lying in K_B.

Consider the lines $\overline{\alpha_1\ \beta_1}$ and $\overline{\beta_j\ \alpha_{j+1}}$. They cannot be parallel, for if they were then $\alpha_1\ \beta_1\ \beta_j\ \alpha_{j+1}$ would be a rectangle whereas we had assumed that the distances between eigenvalues are all distinct. So these lines meet. If the point a at which they meet is closer to α_1 than to β_1 then we can find a disk $\overline{D}(a,\rho)$ containing K_A and a disk $\overline{D}(a,\rho+\delta)$ which contains no point of K_B. So by Proposition 12.1, $\delta \leq \varepsilon$. If the point a is closer to β_1 than to α_1 the same argument works with the roles of A and B interchanged. ∎

§18. An inequality of Sunder

Another instance when the eigenvalues of A and B can be naturally ordered for optimal matching is the case when A is Hermitian and B is skew-Hermitian. The following result was proved by Sunder:

THEOREM 18.1 : Let A be a Hermitian matrix and B a skew-Hermitian matrix with their respective eigenvalues numbered so as to satisfy

$$|\alpha_1| \geq \cdots \geq |\alpha_n| \quad \text{and} \quad |\beta_1| \leq \cdots \leq |\beta_n|.$$

Then for $i = 1,2,\ldots,n$

$$|\alpha_i - \beta_i| \leq \|A-B\|.$$

Proof : For any index j consider the eigenspaces of A and B corresponding to their eigenvalues $\{\alpha_1,\ldots,\alpha_j\}$ and $\{\beta_j,\ldots,\beta_n\}$ respectively. Choose a unit vector x in their intersection. Recall $\|X^*\| = \|X\|$ for every matrix X. We have, therefore,

$$\begin{aligned}
\|A-B\|^2 &= \tfrac{1}{2}\{\|A-B\|^2 + \|A+B\|^2\} \\
&\geq \tfrac{1}{2}\{\|(A-B)x\|^2 + \|(A+B)x\|^2\} \\
&= \|Ax\|^2 + \|Bx\|^2 \\
&\geq |\alpha_j|^2 + |\beta_j|^2 = |\alpha_j - \beta_j|^2.
\end{aligned}$$

(We used the parallelogram identity for the Euclidean vector norm and the fact that α_j is real and β_j imaginary). ■

In section 14 we saw that the inequality $|||(\text{Eig } A, \text{Eig } B)||| \leq |||A-B|||$ does not hold in all unitarily-invariant norms if A,B are unitary. Nor does it hold if A is Hermitian and B skew-Hermitian. To see this let

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

and choose any Schatten p norm for $1 \leq p < 2$. This simple example was also discovered by Sunder.

Notes and references for Chapter 4

One of the most interesting problems concerning spectral variation is the long-standing conjecture that the inequality

$$\|(\text{Eig } A, \text{Eig } B)\| \leq \|A-B\| \tag{1}$$

proved by Weyl for Hermitian matrices would remain true for normal matrices A and B. This conjecture remains open.

However, the stronger conjecture that the Lidskii-Wielandt inequality

$$|||(\text{Eig } A, \text{Eig } B)||| \leq |||A-B||| \tag{2}$$

would also extend from Hermitian to normal matrices turns out to be false, as seen in this chapter, even when A,B are unitary. This stronger conjecture seems to have been first raised explicitly in 1960 by Mirsky [1] and is stated as an open problem in the 1979 book of Marshall and Olkin [1].

That the inequality (1) remains valid for unitary matrices was first proved by Bhatia and Davis [1] and soon afterwards the same result was proved by a different method (and generalized) by Bhatia and Holbrook [1]. Section 17 and 13 are based, respectively, on these two papers. The beautiful Example 13.8 was discovered by Choi [unpublished].

Section 14 is based essentially on the paper of Bhatia [1], where the results upto Theorem 14.6 are proved, and the paper by Bhatia, Davis and McIntosh [1], where Theorem 14.7 and Example 14.8 are given. For the differential geometric notions used here see Auslander and MacKenzie [1]. Proposition 14.1 has been used by Arnol'd [1] for finding a smooth canonical form for matrices under the action of the group GL(n). Theorem 14.9 is common knowledge. (For its history and related matters see Halmos [2]). The proof given here is due to Sunder.

Section 15 is based on a famous paper by Hoffman and Wielandt [1].

The contents of Section 16 are taken from the paper by Bhatia, Davis and McIntosh [1]. The equation (16.1) is known as Rosenblum's Equation (see, fo example, Radjavi and Rosenthal [1]). Theorem 16.1 (in a much more general setting) was proved by Rosenblum [1]. For arbitrary A, B Rosenblum wrote a solution in a form different from (16.4). If Γ is any contour with winding number 1 around every point of Spec (B) and winding number 0 around every point of Spec (A) then Rosenblum's solution to equation (16.1) is given by

$$Q = \frac{-1}{2\pi i} \int_{\Gamma} (A-\zeta)^{-1} S(B-\zeta)^{-1} d\zeta.$$

Without any assumption of normality on A and B no estimate of the type (16.7 can be derived. An example to that effect is given in the above paper of Bhatia, Davis and McIntosh where Theorems 16.2 to 16.6 were first proved. Bounds of the kind (16.10) are called "Sin ϑ theorems". They were studied, among several other things, by Davis and Kahan [1] for the case of Hermitian A, B and for some special kinds of sets K_A, K_B. For an interpretation of $\|EF\|$ in terms of angles between the subspaces E and F see the above paper of Davis and Kahan. (The whole question of perturbation of eigenvectors is discussed in detail in this paper and the paper of Bhatia, Davis and McIntosh).

One proof of the existence of a function f in $L_1(\mathbb{R}^2)$ whose Fourier transform satisfies the relation $\hat{f}(t) = \frac{1}{t_1 + it_2}$ if $|t| \geq 1$ is given in the paper of Bhatia, Davis and McIntosh. Another quick simple proof (due to

M.S. Narasimhan) goes as follows. Let φ be a function on the z-plane which is C^{∞} everywhere, vanishes in a neighbourhood of 0 and is 1 outside the unit disk. Let $\psi(z) = \frac{\varphi(z)}{z}$. We wish to show that $\hat{\psi} \in L_1$. Put $\eta(z) = \frac{d}{d\bar{z}} \psi(z) = \frac{1}{z} \frac{d\varphi}{d\bar{z}}$. (Here, as usual, $\frac{d}{d\bar{z}} = \frac{d}{dx} + i \frac{d}{dy}$). The function

η is a C^{∞} function with compact support. So it belongs to the Schwartz space S. Hence its Fourier transform $\hat{\eta}$ also belongs to S which is contained in L_1. Hence $\hat{\psi}(w) = \frac{\hat{\eta}(w)}{w}$ also belongs to L_1. (We have dropped, in the above discussion, constants like 2π etc. which do not affect the conclusion).

Since such functions exist we can be sure that the constant c_2 is not infinite!

The result in Section 18 was proved by Sunder [2].

The contents of Section 12 are "folklore". Many of them are known with several different proofs and have been used by various authors. Thus, for example, Theorem 12.4 can be regarded as a special case of a theorem of Bauer and Fike [1] which they proved for the case when A is diagonalizable (by a similarity transformation). Another simple proof of Theorem 12.4 may be found in Bhatia and Holbrook [1] (and in several other papers).

5 The general spectral variation problem

In this chapter, we will obtain bounds for the distance d(Eig A, Eig B) when A and B are any two arbitrary matrices.

§19. The distance between roots of polynomials

Consider two monic polynomials

$$f(z) = z^n + a_1 z^{n-1} + \dots + a_n$$

$$g(z) = z^n + b_1 z^{n-1} + \dots + b_n \tag{19.1}$$

with complex coefficients. Let $\alpha_1, \dots, \alpha_n$ and $\beta_1, \dots, \beta_n$ be their respective roots. We will denote by Root f the unordered n-tuple $\{\alpha_1, \dots, \alpha_n$ and also the subset of the plane whose elements are roots of f. (It will be clear from the context which object we are referring to at a particular moment). Let

$$\gamma = \max_{1 \leq k \leq n} (|\alpha_k|, |\beta_k|), \tag{19.2}$$

$$\vartheta = \{\Sigma |b_k - a_k| \gamma^{n-k}\}^{1/n}. \tag{19.3}$$

THEOREM 19.1 : The Hausdorff distance between the root sets of f and g is bounded as

$$h(\text{Root } f, \text{Root } g) \leq \vartheta. \tag{19.4}$$

Proof : We have

$$g(z) - f(z) = \sum_{k=1}^{n} (b_k - a_k) z^{n-k}.$$

So if α_0 denotes any of the roots $\alpha_1,\ldots,\alpha_n$, we have

$$|g(\alpha_0)| \leq \sum_{k=1}^{n} |b_k - a_k|\gamma^{n-k} = \vartheta^n.$$

Factorizing g we can write this as

$$\prod_{i=1}^{n} |\alpha_0 - \beta_i| \leq \vartheta^n \tag{19.5}$$

So, for at least one i we must have $|\alpha_0 - \beta_i| \leq \vartheta$.

Thus if α_j is any root of f, then in the disk $\overline{D}(\alpha_j,\vartheta)$ there is at least one β_i. By symmetry, for every j the disk $\overline{D}(\beta_j,\vartheta)$ contains at least one α_i. This proves the theorem. ■

REMARK 19.2 : The quantity ϑ involves γ and hence depends on the α_i and β_i. However, by an elementary inequality for polynomials, $\max_j |\alpha_j| \leq 2 \max_j |a_j|^{1/j}$. So we can replace γ by 2Γ where $\Gamma = \max_j(|a_j|^{1/j}, |b_j|^{1/j})$.

We now obtain a bound for the optimal matching distance d(Root f, Root g). For this, we will need:

LEMMA 19.3 : Let B be a closed region in the complex plane whose boundary Γ consists of a finite number of circular arcs. Let f and h be two analytic functions on B. Suppose for $0 \leq t \leq 1$, $f(z) + th(z)$ has no zero on Γ. Then the number $N(t)$ of zeros of $f + th$ inside B is constant for $0 \leq t \leq 1$.

Proof : Put $u_t(z) = f(z) + th(z)$. Let t_o be any point in $[0,1]$. By the hypothesis $|u_{t_o}(z)| \geq p > 0$ for all $z \in \Gamma$ and for some p. Choose $|\delta|$ small enough so that $|\delta h(z)| < p$ for all $z \in \Gamma$. Then $|u_{t_o}(z) + \delta h(z) - u_{t_o}(z)| < |u_{t_o}(z)|$ for all $z \in \Gamma$. So, by Rouché's Theorem $u_{t_o} + \delta h$ and u_{t_o} have the same number of zeros inside B. In other words $N(t_o + \delta) = N(t_o)$ for $|\delta|$ sufficiently small. Thus $N(t)$ is a continuous function of t, and being integer-valued it must be constant. ■

THEOREM 19.4 : The optimal matching distance between the n-tuples Root f and Root g is bounded as

$$d(\text{Root } f, \text{Root } g) \leq c(n)\vartheta, \quad (19.6)$$

where $c(n) = n$ or $n-1$ according to whether n is odd or even.

Proof : If $n = 2$ then this follows from Theorem 19.1. So, let $n \geq 3$. Suppose for some α_o picked from $\alpha_1,\ldots,\alpha_n$ we have $|\alpha_o - \beta_i| \geq \vartheta$ for all i. Then from (19.5) we must have $|\alpha_o - \beta_i| = \vartheta$ for all i. So for all i,j we have $|\beta_i - \beta_j| \leq 2\vartheta$. Since every ϑ_k is within a distance ϑ of some β_i, this means that $|\alpha_k - \beta_j| \leq 3\vartheta$ for all k,j. So the theorem is proved in this case.

So, for the remaining part, assume that for every j the *open* disk $D(\alpha_j,\vartheta)$ contains at least one β_i and *vice versa*.

Let $B = \bigcup_{j=1}^{n} D(\alpha_j,\vartheta)$. Let M be a connected component of B. Suppose M is made up of k of these disks. For $0 \leq t \leq 1$ let $g_t(z) = f(z) + t(g(z)-f(z))$

Suppose ξ is any root of $g_1(z)$. Note $g_1 = g$. So by the preceding paragraph $|\xi - \alpha_j| < \vartheta$ for at least one j. Suppose ξ is a root of $g_t(z)$ for some $t < 1$. Then by Theorem 19.1 there exists some j for which $|\xi - \alpha_j| \leq t^{1/n} \vartheta < \vartheta$. So for every $0 \leq t \leq 1$ any root ξ of g_t lies within the open disk $D(\alpha_j, \vartheta)$ for some α_j. Hence, no root of any g_t can lie on the boundary Γ of M. So by Lemma 19.3 each g_t, $0 \leq t \leq 1$ has the same number of roots inside M. Since $g_0 = f$, which has exactly k roots inside M, g_1 also has k roots inside M. But $g_1 = g$. Thus we have shown that every connected component of B contains an equal number of roots of f and g. By symmetry every connected component of the set $\bigcup_{j=1}^{n} D(\beta_j, \beta)$ also contains the same number of α's and β's. So the assertion follows from Theorem 1.5. ∎

We will call an n-tuple $(x_1, \ldots, x_n)$ of complex numbers *Carrollian* if it satisfies the condition : x is a member of this n-tuple iff -x is also a member with the same multiplicity as that of x. Note that when n is odd, a Carrollian n-tuple must contain 0 with an odd multiplicity.

Recall that if $\alpha_1, \ldots, \alpha_n$ are the roots of $f(z) = z^n + a_1 z^{n-1} + \ldots + a_n$ then we have

$$a_k = (-1)^k \, s_k(\alpha_1, \ldots, \alpha_n), \qquad 1 \leq k \leq n$$

where s_k is the elementary symmetric function defined as

$$s_k = \sum_{1 \leq i_1 < \ldots < i_k \leq n} \alpha_{i_1} \cdots \alpha_{i_k} .$$

Now suppose $(\alpha_1, \ldots, \alpha_n)$ is a Carrollian n-tuple. Then we can arrange it as $(\alpha_1, \ldots, \alpha_r, -\alpha_1, \ldots, -\alpha_r)$ when $n = 2r$ and as $(0, \alpha_1, \ldots, \alpha_r, -\alpha_1, \ldots, -\alpha_r)$ when $n = 2r+1$. In this case we have

$$a_{2j} = (-1)^j \, s_j(\alpha_1^2, \ldots, \alpha_r^2) \qquad 1 \leq j \leq r,$$

$$a_{2j+1} = 0 \qquad 0 \leq j \leq r.$$

Thus $\alpha_1^2,\ldots,\alpha_r^2$ are the r roots of the polynomial

$$F(z) = z^r + a_2 z^{r-1} +\ldots+ a_{2j} z^{r-j} +\ldots+ a_{2r}.$$

This observation leads to the following corollary of Theorem 19.4.

COROLLARY 19.5 : Let f,g by polynomials defined by (19.1). Let $n = 2r$ or $2r+1$. Suppose the roots $\alpha_1,\ldots,\alpha_n$ of f(z) and $\beta_1,\ldots,\beta_n$ of g(z) both form Carrollian n-tuples. Let

$$\vartheta' = \{\sum_{k=1}^{r} |b_{2k} - a_{2k}| \gamma^{2(r-k)}\}^{1/r}, \tag{19.7}$$

where γ is defined by (19.2). Then the roots can be arranged in such a way that for $j = 1,2,\ldots,n$

$$|\alpha_j^2 - \beta_j^2| \le c(r)\vartheta' \tag{19.8}$$

where $c(r) = r$ or $r - 1$ according to whether r is odd or even.

Note that (19.8) gives a bound for the optimal matching distance between the squares of the roots of f and g. A bound for the optimal matching distance between the roots themselves can be obtained under an additional hypothesis.

COROLLARY 19.6 : Let notations be as in Corollary 19.5. Suppose, in addition, that either the roots $\alpha_1,\ldots,\alpha_n$ and $\beta_1,\ldots,\beta_n$ are all located outside a circle of radius R around the origin, or they contain 0 with the same multiplicity and the nonzero ones are located outside this circle. Then

$$d(\text{Root } f, \text{ Root } g) \le \frac{c(r)}{R} \vartheta'. \qquad (19.9)$$

Proof : Use the fact that if x_1, x_2 are any two complex numbers, each of modulus not less than R, then either $|x_1 - x_2| \ge R$ or $|x_1 + x_2| \ge R$. Then (19.9) follows from (19.8). ■

§20. Variation of Grassman powers and spectra

One strategy for deriving spectral variation bounds could be: first study the variation of the coefficients of the characteristic polynomial of a matrix and then use the results of section 19 to get bounds on d(Eig A, Eig B). As pointed out in section 4, if the characteristic polynomial of a matrix A is written as

$$\chi_A(t) = t^n - a_1 t^{n-1} + a_2 t^{n-2} + \ldots + (-1)^n a_n,$$

then

$$a_k = \operatorname{tr} \Lambda^k A, \qquad k = 1,2,\ldots,n,$$

where $\Lambda^k A$ denotes the kth Grassman power of A. So, to study the variation of a_k, we should begin by studying the variation of the map $A \to \Lambda^k A$. This is best done in the language of calculus in Banach spaces. (See, for example, Dieudonné [1]).

For a Hilbert space H let us denote by $L(H)$ the Banach space of all linear operators on H equipped with the operator norm. Consider the kth tensor power $\otimes^k H$ of H and the induced map $\otimes^k : A \to \otimes^k A$ from $L(H)$ to $L(\otimes^k H)$. (See section 4 for notations.) The *derivative* of $\otimes^k$ at a point A is the linear map $D\otimes^k(A)$ from $L(H)$ to $L(\otimes^k H)$ defined as

$$D\otimes^k(A)(B) = \frac{d}{dt}\bigg|_{t=0} \otimes^k (A + tB)$$

(the directional derivative of $\otimes^k$ at A in the direction of B). By the multilinearity properties of tensor powers one sees
$D\otimes^k (A)(B) = B\otimes A\otimes \dots \otimes A + A\otimes B\otimes \dots \otimes A + \dots + A\otimes \dots \otimes A \otimes B$.
Since $\|X \otimes Y\| = \|X\| \, \|Y\|$ for any two operators X,Y, this shows

$$\|D\otimes^k(A)(B)\| \leq k\|A\|^{k-1}\|B\|.$$

Taking supremum over all B with $\|B\| = 1$ and considering the special case $B = A/\|A\|$ we obtain

$$\|D\otimes^k (A)\| = k\|A\|^{k-1} . \qquad (20.1)$$

Now let $P_k : \otimes^k H \to \Lambda^k H$ be the canonical projection operator defined in section 4 and let $Q_k : \Lambda^k H \to \otimes^k H$ be the inclusion map. Consider the induced map $\tilde{P}_k : L(\otimes^k H) \to L(\Lambda^k H)$ defined by

$$\tilde{P}_k(T) = P_k T Q_k \quad \text{for all} \quad T \in L(\otimes^k H).$$

Then $\tilde{P}_k$ is a projection and $\|\tilde{P}_k\| = 1$. The map $\Lambda^k : L(H) \to L(\Lambda^k H)$ factors through the map $\otimes^k : L(H) \to L(\otimes^k H)$ via the projection $\tilde{P}_k$, i.e.

$$\Lambda^k(A) = \tilde{P}_k(\otimes^k A) \quad \text{for all} \quad A \in L(H).$$

By the chain rule of differentiation and the fact that the derivative of the linear map $\tilde{P}_k$ is $\tilde{P}_k$ itself, we have

$$D\Lambda^k(A) = \tilde{P}_k \cdot D(\otimes^k A). \qquad (20.2)$$

From (20.1) and (20.2) we have

THEOREM 20.1 : For $k = 1,2,\ldots,n$

$$\|D\Lambda^k(A)\| \le k\|A\|^{k-1} . \tag{20.3}$$

COROLLARY 20.2 : For any A, B in $L(H)$ we have

$$\|\Lambda^k B - \Lambda^k A\| \le k\, M^{k-1}\|B-A\|. \tag{20.4}$$

where

$$M = \max(\|A\|,\|B\|).$$

Proof : Consider the map $f : [0,1] \to L(H)$ defined by $f(t) = (1-t)A + tB$. Apply the mean value theorem to the composite map $\Lambda^k \circ f$ from $[0,1]$ to $L(\Lambda^k H)$. This gives

$$\|\Lambda^k B - \Lambda^k A\| \le \sup_{0\le t\le 1} \|D\Lambda^k f(t)\|\, \|Df(t)\|$$

$$\le \sup_{0\le t\le 1} k\|(1-t)A + tB\|^{k-1}\, \|B-A\|$$

$$\le k\, M^{k-1}\, \|B-A\| \quad ■$$

Denote by $(\mathrm{tr})_n$ the trace map from the space $L(H)$ to $\mathbb{C}$, where $\dim H = n$. This map is linear and hence

$$\|D(\mathrm{tr})_n(A)\| = \|(\mathrm{tr})_n\| = n \quad \text{for all} \quad A \in L(H).$$

Since dim $\Lambda^k H = \binom{n}{k}$, using Theorem 20.1 and the chain rule of differentiation we get

PROPOSITION 20.3 : Let $\varphi_k(A) = \operatorname{tr} \Lambda^k A$ be the kth coefficient in the characteristic polynomial of A. Then

$$\|D\varphi_k(A)\| \leq k \binom{n}{k} \|A\|^{k-1}. \tag{20.5}$$

If A, B are any two operators on H, then

$$|\varphi_k(B) - \varphi_k(A)| \leq k \binom{n}{k} M^{k-1} \|B-A\|, \tag{20.6}$$

where $M = \max(\|A\|,\|B\|)$.

Using this and results of section 19 we obtain

THEOREM 20.4 : Let A, B be any two n×n matrices. Then the Hausdorff distance between their spectra and the optimal matching distance between their eigenvalues are bounded as

$$h(\operatorname{Spec} A, \operatorname{Spec} B) \leq n^{1/n}(2M)^{1-1/n} \|A-B\|^{1/n}, \tag{20.7}$$

$$d(\operatorname{Eig} A, \operatorname{Eig} B) \leq c(n)\, n^{1/n} (2M)^{1-1/n} \|A-B\|^{1/n}, \tag{20.8}$$

where $M = \max(\|A\|,\|B\|)$ and where $c(n) = n$ or $n-1$ according to whether n is odd or even.

Proof : Use Theorem 19.1 and Proposition 20.3. Use the fact that the eigenvalues of A are bounded in modulus by $\|A\|$. This gives

$$h(\text{Spec } A, \text{ Spec } B) \leq \left\{ \sum_{k=1}^{n} |\varphi_k(B) - \varphi_k(A)| M^{n-k} \right\}^{1/n}$$

$$\leq \left\{ \sum_{k=1}^{n} k \binom{n}{k} M^{n-1} \|B-A\| \right\}^{1/n}.$$

Now use the combinatorial identity

$$\sum_{k=0}^{n} k \binom{n}{k} = n\, 2^{n-1}$$

to get (20.7). The inequality (20.8) follows likewise from Theorem 19.4. ■

REMARK 20.5 : The bounds given by Theorem 20.4 differ from those derived for special matrices in Chapters 3 and 4. These earlier results bounded the spectral distance by expressions like $c\|A-B\|$ where c was a constant independent of A, B and n. In most cases c was 1. Now the bounds (20.7) and (20.8) involve $\|A-B\|^{1/n}$, a common bound M for $\|A\|$, $\|B\|$ and a constant growing with n. Let us see whether this is the best we can do.

First, let A be the nilpotent upper Jordan matrix of order n (i.e. A has ones on the first super diagonal and zeros elsewhere). Let B be the matrix obtained by adding an entry ε in the southwest corner of A. Then $\|A-B\| = \varepsilon$. But the eigenvalues of A are all zero and the eigenvalues of B are the nth roots of ε. So the left-hand sides of (20.7) and (20.8) are both equal to $\varepsilon^{1/n}$. Thus the order 1/n with which $\|A-B\|$ occurs in these estimates cannot be improved in general. Note that, in perturbation theory, we are mainly interested in the case when A is close to B. Then $\|A-B\|^{1/n}$ is much larger than $\|A-B\|$.

Next, note that if we multiply A and B by a constant t then the left-hand sides of (20.7) and (20.8) will be multiplied by t. But $\|A-B\|^{1/n}$ will be multiplied by only $t^{1/n}$. So, if the right-hand side has to involve $\|A-B\|^{1/n}$ then it must involve a compensatory factor $M^{1-1/n}$.

Now we have to see whether the remaining factor $n^{1/n}.2^{1-1/n}$ in (20.7) can be improved. Take A = I and B = -I. Then h(Spec A, Spec B) = 2. But $M^{1-1/n}\ \|A-B\|^{1/n} = 2^{1/n}$. So, the factor by which this should be multiplied to bound h(Spec A, Spec B) has to grow with n. Let $\varphi(n)$ be the optimal choice for this factor, i.e. $\varphi(n)$ is the smallest constant depending on n such that

$$h(\text{Spec } A, \text{Spec } B) \leq \varphi(n)\ M^{1-1/n}\ \|A-B\|^{1/n}.$$

Then by the above example and by (20.7)

$$2^{1-1/n} \leq \varphi(n) \leq n^{1/n}\ 2^{1-1/n}.$$

So our bound (20.7) is quite sharp.

The bound (20.8) involves an additional factor c(n) = n or n-1. It is possible that this could be drastically cut down.

§21. Some more spectral variation bounds

In this section, we will deviate from our practice of using, in every spectral variation inequality, the same norm to measure the distance $|||(\text{Eig } A, \text{Eig } B)|||$ between eigenvalues and the distance $|||A-B|||$ between operators. This has served us well so far. But there are some results known for arbitrary matrices which give bounds for d(Eig A, Eig B) in terms of norms other than the operator norm. We will summarize here these results without proof.

Given an nxn matrix $A = ((a_{ij}))$ define its L-norm as

$$\|A\|_L = \frac{1}{n} \sum_{i,j} |a_{ij}| .$$

(This norm is easy to compute for a matrix. But it is not unitarily-invariant, so it is defined only for a particular matrix representation of an operator).

<u>THEOREM 21.1</u> : For any two nxn matrices A, B

$$d(\text{Eig } A, \text{Eig } B) \leq c(n)(n+2)\, M_L^{1-1/n}\, \|A-B\|_L^{1/n} \qquad (21.1)$$

where $M_L = \max(\|A\|_L, \|B\|_L)$ and where $c(n) = n$ or $n-1$ according to whether n is odd or even.

The Frobenius norm $\|A\|_F = \left(\sum_{i,j} |a_{ij}|^2 \right)^{1/2}$ is almost as easy to calculate for every matrix as the norm $\|A\|_L$. It has the additional advantage of being unitarily-invariant. For this norm we have:

<u>THEOREM 21.2</u> : For any two nxn matrices A, B

$$d(\text{Eig } A, \text{Eig } B) \leq c(n) \left\{ \sum_{k=1}^{n} k^{1-k/2} \binom{n}{k} \right\}^{1/n} M_F^{1-1/n}\, \|A-B\|_F^{1/n} \qquad (21.2)$$

where $M_F = \max(\|A\|_F, \|B\|_F)$ and where $c(n) = n$ or $n-1$ according to whether n is odd or even.

Note that

$$\sum_{k=1}^{n} k^{1-k/2} \binom{n}{k} \leq \sum_{k=1}^{n} \binom{n}{k} = 2^n - 1 \tag{21.3}$$

and the inequality is strict for n > 2. Using this (21.2) can be replaced by a weaker but neater inequality

$$d(\text{Eig } A, \text{ Eig } B) \leq 2\ c(n)\ M_F^{1-1/n}\ \|A-B\|_F^{1/n}\ . \tag{21.4}$$

REMARK 21.3 : The following relations between norms can be easily verified

$$\|A\| \leq \|A\|_F \leq \sqrt{n}\ \|A\|$$

$$\|A\|_L \leq \|A\|_F \leq n\|A\|_L .$$

Using these relations it is not possible to derive either of the above theorems from the other, or from Theorem 20.4.

However, note that using these relations we get from (21.2)

$$d(\text{Eig } A, \text{ Eig } B) \leq c(n)\ (2^n - 1)^{1/n}\ n\ M_L^{1-1/n}\ \|A-B\|_L^{1/n}\ . \tag{21.5}$$

So, after this crude substitution the inequality (21.5) is still better than (21.1) for n = 2, by a factor of $\sqrt{3}/2$. For n > 2 it is worse than (21.1) by a factor smaller than $\frac{2n}{n+2}$, which is less than 2. However, after a similar substitution the inequality (21.1) suffers in comparison with (21.4) by a factor of $\frac{n+2}{2}$.

Another approach to the study of spectral variation was followed by Henrici, who instead of using the characteristic polynomial employed some estimates of the norm of the resolvent $(A-\mu I)^{-1}$ and a measure of non-normality of a matrix.

Given a matrix A we can find a unitary matrix U such that

$$UAU^* = T = D + N$$

where T is an upper triangular matrix, D and N are the diagonal and the nilpotent parts of T. Such a T is not unique. If ν is any matrix norm, define the ν-*departure from normality* of A by

$$\Delta_\nu(A) = \inf \nu(N)$$

where the infimum is taken over all N which appear as the nilpotent parts in the upper triangular forms of A. Note that $\Delta_\nu(A) = 0$ iff A is normal.

The measure of nonnormality is difficult to evaluate. For the Frobenius norm one can prove that

$$\Delta_F(A) \leq \left(\frac{n^3 - n}{12}\right)^{1/4} \|A^*A - AA^*\|_F^{1/2} . \tag{21.6}$$

This inequality and the following theorem were proved by Henrici:

<u>THEOREM 21.4</u> : Let A be a nonnormal matrix and let $B \neq A$. Let ν by any matrix norm which majorizes the operator norm. Let

$$y = \frac{\Delta_\nu(A)}{\nu(B-A)} \tag{21.7}$$

and let g(y) be the unique nonnegative solution of the equation

$$g + g^2 + \ldots + g^n = y.$$

Let

$$\eta = \frac{y}{g(y)} \, \dot{\nu}(A-B) \ . \tag{21.8}$$

Then every eigenvalue of B is within a distance η of an eigenvalue of A. Further

$$d(\text{Eig } A, \text{ Eig } B) \leq (2n - 1)\eta. \tag{21.9}$$

If A is normal then the expression $y/g(y)$ in (21.8) can be replaced by 1.

REMARKS : When A is normal the first part of this theorem reduces to Theorem 12.4.

However, the bounds for d(Eig A, Eig B) obtained earlier for the case when both A and B are normal cannot be deduced from (21.9).

The bound (21.9) has two *apparent* weaknesses when compared to other bounds presented in this chapter (the inequalities (20.8), (21.1) and (21.4)). It involves more complicated expressions and it is "local" in the sense that it uses some special knowledge about one of the matrices A via (21.7). However, Elsner has shown that by a little modification of Henrici's argument, together with some inequality manipulations, one can derive not only the bounds (20.8) and (21.2) but also slight improvements thereof. Thus for example, using this method Elsner shows that the factor $n^{1/n}$ occurring in (20.8) can be replaced by

$$\gamma_n = \left(1 + \frac{1}{c(n)} + \ldots + \frac{1}{c(n)^{n-1}}\right)^{1/n} .$$

For $n > 2$ we have

$$\gamma_n < \left(1 + \frac{1}{n-2}\right)^{1/n} < n^{1/n} .$$

§22. Spectral variation for the classical Lie algebras

Let A^t denote the transpose of the matrix A. We call a complex matrix A *symmetric* if $A^t = A$ and *skew-symmetric* if $A^t = -A$. Let I_n denote the nxn identity matrix and let J denote a (2r) × (2r) matrix with a block decomposition

$$J = \begin{bmatrix} 0 & I_r \\ I_r & 0 \end{bmatrix}.$$

Let

$$\mathfrak{so}(n,\mathbb{C}) = \{\text{nxn complex skew-symmetric matrices}\}$$

$$\mathfrak{sp}(r,\mathbb{C}) = \{A : A^t = -JAJ^{-1}\}.$$

It is easy to see that $A \in \mathfrak{sp}(r,\mathbb{C})$ if

$$A = \begin{bmatrix} A_1 & A_2 \\ A_3 & -A_1^t \end{bmatrix}$$

where A_1, A_2, A_3 are rxr matrices of which A_2, A_3 are skew-symmetric.

It is clear that the eigenvalues of matrices coming from these two sets form Carrollian n-tuples. So we can use the method of section 20 together with Corollary 19.5 to obtain different spectral variation bounds for such matrices.

We thus have:

THEOREM 22.1 : Let A,B be two matrices of order n = 2r or n = 2r+1. Suppose A,B are elements either of the Lie algebra $\mathfrak{so}(n,\mathbb{C})$ or the Lie algebra $\mathfrak{sp}(r,\mathbb{C})$. Then

$$d(\mathrm{Eig}\ A^2, \mathrm{Eig}\ B^2) \leq 2^{(n-2)/r}\ c(r)\ n^{1/r}\ M^{2-1/r}\ \|A-B\|^{1/r} \tag{22.1}$$

where $M = \max(\|A\|, \|B\|)$ and $c(r)$ is r or $r-1$ depending on whether r is odd or even.

Further, if A, B both have 0 as one of their eigenvalues with the same multiplicity and the rest of their eigenvalues lie outside a circle or radius R around the origin, then we have

$$d(\mathrm{Eig}\ A, \mathrm{Eig}\ B) \leq w/R \tag{22.2}$$

where w denotes the right hand side of (22.1).

The point of these inequalities is that the order $1/n$ with which $\|A-B\|$ entered on the right hand side of (20.8) has now been improved to $1/r$.

Notes and references for Chapter 5

The results and the methods of section 19 are essentially due to Ostrowski [1], [3]. Theorem 19.1 was proved by him. Theorem 19.4 was proved by him in a weaker form, with a factor of $(2n-1)$ in place of $c(n)$ in (19.6). The significant improvement to the constant $c(n)$ in place of $(2n-1)$ is due to Elsner [1]. The crucial observation of Elsner is that, exploiting symmetry, one can use Theorem 1.5 for matching the roots rather than Theorem 1.6 which Ostrowski used. Ostrowski gave an example to show that the bound (19.4) cannot be improved and another example to show that the factor $c(n)$ in (19.6) cannot be replaced by 1.

It has been well known to numerical analysts that even though the roots of a polynomial vary continuously with the coefficients they can change drastically with small perturbations of the coefficients. Wilkinson [2] gave the following example which caused considerable consternation among numerical analysts. Let

$$f(x) = \prod_{j=1}^{20} (x + j) = x^{20} + 210\ x^{19} + \ldots + 20!.$$

The zeros of this polynomial are -20, -19,...,-1. If the coefficient 210 of x^{19} is changed to $210 + 2^{-23}$ the new polynomial has 10 real roots -20, -9, -8,...,-1 and 10 complex roots with large imaginary parts. Nevertheless, the change in the roots can be bounded by a function of the change in coefficients as shown in Section 19. For an interesting discussion of Wilkinson's example see Poston and Stewart [1].

An interesting formulation of the continuity of the roots of a polynomial is the following topological theorem. Let $\mathbb{C}^n_{sym} = \mathbb{C}^n/S_n$ be the space of unordered n-tuples of complex numbers, obtained as a quotient space under the action of the permutation group S_n on $\mathbb{C}^n$. Then the map $S : \mathbb{C}^n_{sym} \to \mathbb{C}^n$ which takes $\{x_1,\dots,x_n\}$ to $(s_1,\dots,s_n)$, where s_j is the jth elementary symmetric polynomial in the variables $x_1,\dots,x_n$, is a homeomorphism. Recall that $s_1,\dots,s_n$ are the coefficients of a monic polynomial of degree n which has $x_1,\dots,x_n$ as roots. So the continuity of the roots follows from this homeomorphism. For an easily accessible proof of this well-known theorem see Bhatia and Mukherjea [2].

The idea of studying spectral variation via calculus and exterior algebra occurs in a paper of Bhatia and Mukherjea [1]. Using this a weaker version of Theorem 21.2 was proved there. The same approach was followed, with greater success, by Bhatia and Friedland [1]. Section 20 is based on this paper. To us this approach seems appealing because $A \to \text{Eig } A$ is a map from the space M(n) of matrices into the space $\mathbb{C}^n_{sym}$. If S is the homeomorphism mentioned in the preceding paragraph then the composite map $S \circ \text{Eig}$ takes a matrix A to the n tuple $(a_1,\dots,a_n)$ where $a_k = \text{tr } \Lambda^k A$. So the variation of the map Eig is best studied via the variation of the map $A \to \Lambda^k A$. In the above mentioned paper of Bhatia and Friedland a result much stronger than Theorem 20.1 is proved. It is shown there that

$$\|D\Lambda^k(A)\| = s_{k-1}(\nu_1,\dots,\nu_n)$$

where $\nu_1 \geq \dots \geq \nu_n$ are the singular values of A and s_{k-1} is the (k-1)th elementary symmetric polynomial.

It is interesting to note that the operator $A^{[k]}$ introduced in section 6 is equal to the derivative $D\Lambda^k(I)(A)$.

Theorem 21.1 was proved by Ostrowski [2], [3], with the constant (2n-1) in place of c(n). Theorem 21.4 is due to Henrici [1]. In this paper he proved several other results for arbitrary matrices in terms of their measure of nonnormality. For instance, we know that the convex hull H(A) of the eigenvalues of a matrix A is equal to its numerical range W(A) if A is normal; otherwise H(A) is contained in W(A). One of the theorems in the above paper of Henrici gives a bound for the distance between the boundary of W(A) and the set H(A) in terms of the measure of nonnormality of A.

In an important paper Elsner [1] improved all these spectral variation bounds obtained by Ostrowski, Bhatia-Mukherjea and Bhatia-Friedland by cutting down the factor (2n-1) which originally occurred in these results to c(n). In the same paper he showed how to derive Theorems 20.4 and 21.2 by Henrici's method.

Theorem 22.1, in a weaker form, is proved in Bhatia [2].

Notice that for k = n, the inequality (20.4) gives a perturbation bound for the determinant

$$|\det A - \det B| \leq n\, M^{n-1}\, \|A-B\|.$$

Using the same analysis one can prove that the right-hand side above also dominates $|\text{per } A - \text{per } B|$ where per A is the permanent of A. (See Bhatia [3]).

For an analysis parallel to that in section 20 in some more general norms see Friedland [1].

6 Arbitrary perturbations of constrained matrices

All our results obtained so far had one common feature: the matrices A, B either both belonged to a familiar class simultaneously or were completely arbitrary. (One exception was made in section 18). This is justified because most often the context demands that both the matrices must satisfy the same constraint; for example, symmetric matrices occur most frequently in physical problems. However, when a matrix is subjected to random errors then it may go outside the class to which it initially belonged. In this chapter, we present some results on spectral variation when one of the matrices is Hermitian or normal and the other is arbitrary.

The other feature of our results has been that all the bounds we derived were global and *a priori*. The first adjective refers to the fact that we assumed no special knowledge about any *one* of the matrices save the fact that it belonged to a special class. (One exception was made in section 21 where a measure of nonnormality was required for one of the matrices). By an *a priori* bound we mean a bound which uses no knowledge about the eigenvalues of one of the matrices. In practice such knowledge is often available. We will not study any of the several useful *a posteriori* bounds known. We do present one more local result - the celebrated Bauer-Fike Theorem.

Proofs in this chapter will be somewhat sketchy.

§23. Arbitrary perturbations of Hermitian matrices and Kahan's results

Suppose Z is an nxn complex matrix with real spectrum. Then we have

$$\|Z-Z^*\|_F \leq \|Z+Z^*\|_F. \tag{23.1}$$

(The easiest way of seeing this is by converting Z into an upper triangular form). Since $\|A\| \leq \|A\|_F \leq \sqrt{n}\,\|A\|$, this gives

$$\|Z-Z^*\| \le \sqrt{n}\ \|Z+Z^*\|. \tag{23.2}$$

This is, however, too crude. An interesting refinement was obtained by Kahan.

THEOREM 23.1 : If Z is an nxn complex matrix with real spectrum then

$$\|Z-Z^*\| \le \gamma_n \|Z+Z^*\|, \tag{23.3}$$

where the constant γ_n which depends only on n is bounded as

$$\frac{2}{\pi} \ln n - O(1) \le \gamma_n \le \log_2 n + 0.038. \tag{23.4}$$

This bound for the constant occurring in (23.3) was slightly improved by Schönage. Later Pokrzywa evaluated the best constant exactly. His result can be stated as

THEOREM 23.2 : Let

$$\gamma_n = \max \frac{\|Z-Z^*\|}{\|Z+Z^*\|}, \tag{23.5}$$

where the maximum is taken over all complex nxn matrices Z with real spectrum. Then

$$\gamma_n = \frac{2}{n} \sum_{j=1}^{[n/2]} \cot \frac{2j-1}{2n} \pi, \tag{23.6}$$

where $\left[\frac{n}{2} \right]$ denotes the integral part of $\frac{n}{2}$.

We will assume the above results and use them to prove

THEOREM 23.3 : Let A be a Hermitian matrix with eigenvalues $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ and let B be an arbitrary matrix with eigenvalues $\beta_1, \ldots, \beta_n$ arranged so that $\mathrm{Re}\, \beta_1 \geq \mathrm{Re}\, \beta_2 \geq \ldots \geq \mathrm{Re}\, \beta_n$. Then

$$\max_{1 \leq j \leq n} |\alpha_j - \beta_j| \leq (\gamma_n + 2) \, \|A-B\|, \tag{23.7}$$

where γ_n is the constant defined by (23.6).

Proof : Put $C = B-A$, $X = \frac{C+C^*}{2}$, $Y = \frac{C-C^*}{2i}$. We have $\|X\| \leq \|C\|$, $\|Y\| \leq \|C\|$, $\|C\| \leq \|X\| + \|Y\|$.

Since the eigenvalues and the norms are invariant under unitary conjugation, we can assume, without loss of generality that B is upper triangular and write

$$B = L + iM + iN$$

where $L = \mathrm{diag}(\mathrm{Re}\, \beta_1, \ldots, \mathrm{Re}\, \beta_n)$, $M = \mathrm{diag}(\mathrm{Im}\, \beta_1, \ldots, \mathrm{Im}\, \beta_n)$, and N is a strictly upper triangular matrix. Note that

$$A + X = \frac{B+B^*}{2} = L + i \, \frac{N-N^*}{2} \, ,$$

$$Y = \frac{B-B^*}{2i} = M + \frac{N+N^*}{2} \, .$$

Now note that $\mathrm{Im}\, \beta_j$ are the diagonal entries of the matrix $\frac{B-B^*}{2i}$. Hence

$$|\mathrm{Im}\, \beta_j| \leq \frac{1}{2} \|B-B^*\| = \|Y\|. \tag{23.8}$$

By Weyl's Inequality (Theorem 8.1)

$$|\alpha_j - \text{Re}\ \beta_j| \leq \|A-L\| = \|X - \frac{i}{2}(N - N^*)\|$$

$$\leq \|X\| + \frac{1}{2}\|N-N^*\|.$$

Put $Z = M+N$. Then Z has real spectrum. Further $Z-Z^* = N-N^*$, $Z+Z^* = 2M + N + N^* = 2Y$. So by Theorem 23.1

$$\|N-N^*\| \leq 2\gamma_n \|Y\|.$$

Hence

$$|\alpha_j - \text{Re}\ \beta_j| \leq \|X\| + \gamma_n \|Y\|.$$

So,

$$|\alpha_j - \beta_j| \leq |\alpha_j - \text{Re}\ \beta_j| + |\text{Im}\ \beta_j|$$

$$\leq \|X\| + (\gamma_n + 1)\|Y\|$$

$$\leq (\gamma_n + 2)\|C\|.$$

This proves the theorem. ∎

EXAMPLE 23.4 : Let A be the Hermitian matrix with entries

$$a_{ij} = \frac{1}{|i-j|} \quad \text{if} \quad i \neq j$$

$$a_{ii} = 0, \qquad i = 1,2,\ldots,n.$$

Let C be the skew-Hermitian matrix with entries

$$c_{ij} = \frac{1}{i-j} \quad \text{if} \quad i \neq j$$

$$c_{ii} = 0, \qquad i = 1,2,\ldots,n.$$

Let $B = A+C$. Then B is strictly lower triangular. So Spec $B = \{0\}$. It was shown by Kahan that $\|C\| \leq \pi$ (independent of n), whereas $2 \log n - O(1) < \|A\| < 2 \log n$. Since A is Hermitian this means that the spectral radius of A grows like $\log n$. So $d(\text{Eig } A, \text{Eig } B)$ grows as $\log n$ whereas $\|A-B\| \leq \pi$. Thus the right-hand side of (23.7) must involve a factor like γ_n. Further it can be seen from (23.6) that $\frac{\gamma_n}{\log n} \to \frac{2}{\pi}$ as $n \to \infty$. So, this example shows that the bound (23.7) is not too loose.

REMARK 23.5 : If B is normal then $N = 0$ (in the proof of Theorem 23.3). In this case we have

$$|\alpha_j - \beta_j|^2 = |\alpha_j - \text{Re } \beta_j|^2 + |\text{Im } \beta_j|^2 \leq \|X\|^2 + \|Y\|^2$$
$$\leq 2\|C\|^2 = 2\|A-B\|^2.$$

So we have

$$\max_{1 \leq j \leq n} |\alpha_j - \beta_j| \leq \sqrt{2} \, \|A-B\| \tag{23.9}$$

when A is Hermitian and B is normal

§24. Arbitrary perturbations of normal matrices

THEOREM 24.1 : Let A, B be $n \times n$ matrices. If A is normal then

$$d(\text{Eig } A, \text{Eig } B) \leq (2n-1) \, \|A-B\| \tag{24.1}$$

Proof : Let $\alpha_1,\dots,\alpha_n$ and $\beta_1,\dots,\beta_n$ be the respective eigenvalues of A and B. We have shown (Theorem 12.4) that if $\varepsilon = \|A-B\|$ then

$$\text{Spec } B \subset \bigcup_{j=1}^{n} \overline{D}(\alpha_j,\varepsilon) = D$$

where $\overline{D}(\alpha_j,\varepsilon)$ is the closed disk with centre α_j and radius ε.

Let $A(t) = (1-t)A + tB$, $0 \le t \le 1$. Then $\|A(t) - A\| = t\varepsilon \le \varepsilon$ for all t. So, Spec A(t) is also contained in D for $0 \le t \le 1$. By the argument used in section 19 one can conclude from this that each connected component of D contains as many α's as β's. The theorem now follows from Theorem 1.6. ■

Remarks : If both A, B were given to be normal then this argument, via Theorem 1.5 would lead to the inequality $d(\text{Eig } A, \text{Eig } B) \le c(n) \|A-B\|$ where $c(n) = n$ or $n-1$ according to whether n is odd or even. But in this case we have proved a stronger result in section 16 where we showed that c(n) here could be replaced by a constant independent of n.

Henrici's Theorem - the inequality (21.9) - is a very good generalization of the above theorem. In fact, Henrici's result reduces to (24.1) when A is normal.

It seems likely that the constant $2n-1$ occurring in (24.1) could be reduced.

§25. The Bauer-Fike theorem

Our main concern in these notes has been finding bounds for $\|\|(\text{Eig } A, \text{Eig } B)\|$. We have, however, occasionally run into some results which say that Spec B is contained in the union of certain disks around the eigenvalues of A. (See Theorem 12.4 and Theorem 21.4). We will not attempt to list all such results. Only the two most famous results of this type are given in this section.

If T is an invertible matrix then the *condition number* of T is defined as

$$c(T) = \|T\| \, \|T^{-1}\|.$$

THEOREM 25.1 (Bauer-Fike) : Let A be similar to a diagonal matrix, i.e. suppose there exists an invertible matrix T such that $A = T\Lambda T^{-1}$ where $\Lambda = \mathrm{diag}(\alpha_1,\ldots,\alpha_n)$. Let B be an arbitrary matrix. Let

$$\varepsilon = \|A-B\|\cdot c(T).$$ Then

$$\mathrm{Spec}\, B \subset \bigcup_{i=1}^{n} \overline{D}(\alpha_i,\varepsilon).$$

Proof : Let $\beta \in \mathrm{Spec}\, B$ and choose a nonzero vector x such that $Bx = \beta x$. Suppose $\beta \notin \mathrm{Spec}\, A$. Then the equation $(B-A)x = (\beta-A)x$ can be rewritten as

$$x = (\beta I-A)^{-1}(B-A)x$$

$$= [T(\beta I-\Lambda)T^{-1}]^{-1}(B-A)x$$

$$= T(\beta I-\Lambda)^{-1}\, T^{-1}(B-A)x.$$

So

$$\|x\| \le \|(\beta I-\Lambda)^{-1}\| \, \|B-A\| \, c(T) \, \|x\|$$

i.e. $1 \le \|(\beta I-\Lambda)^{-1}\|\varepsilon$

$$= \max_j |\beta-\alpha_j|^{-1} \varepsilon$$

i.e. $\min_j |\beta-\alpha_j| \le \varepsilon.$

This proves the Theorem. ■

Remarks : (1) The properties of norms used in the above proof are

$$\|Ax\| \leq \|A\| \, \|x\|, \qquad \text{for all vectors } x$$

$$\|AB\| \leq \|A\| \, \|B\|,$$

$$\|\Lambda\| = \max |\alpha_i| \quad \text{if } \Lambda = \mathrm{diag}(\alpha_1,\ldots,\alpha_n).$$

If $\|x\|$ is any norm defined on $\mathbb{C}^n$, then a norm defined on matrices which satisfies the above conditions is said to be a matrix norm consistent with the given vector norm $\|x\|$.

(2) When A is normal then the invertible matrix T is unitary. So $c(T) = 1$. So we get Theorem 12.4 in this special case.

(3) Using the continuity argument employed in section 19 and also in the proof of Theorem 24.1, we can see that if D is any connected component of the union of the n disks of the Theorem then D contains as many eigenvalues of B as of A.

In a similar vein we have:

THEOREM 25.2 : Let B be any matrix with entries b_{ij}. Then the eigenvalues of B lie in the union of the n *Gersgorin disks* $\{z : |z-b_{ii}| \leq \sum_{j\neq i} |b_{ij}|\}$, $i = 1,2,\ldots,n$.

Proof : Imitate the above proof of the Bauer-Fike Theorem using the vector norm $\|x\|_\infty = \max_{1\leq i\leq n} |x_i|$ on $\mathbb{C}^n$ and the corresponding operator norm

$$\|A\|_\infty = \sup\{\|Ax\|_\infty : \|x\|_\infty = 1\}$$

$$= \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|.$$

This norm satisfies the conditions of Remark (1). Note that if $\|A\|_\infty < 1$ then I-A is invertible.

Let $\Lambda = \operatorname{diag}(b_{11},\ldots,b_{nn})$ and $H = B-\Lambda$. Let β be an eigenvalue of B and suppose $\beta \neq b_{ii}$ for any i. Then

$$\beta I-B = \beta I - H - \Lambda = (\beta I-\Lambda)\{I - (\beta I-\Lambda)^{-1} H\}.$$

Since $\beta I-\Lambda$ is invertible and $\beta I-B$ is not, we must have

$$1 \leq \|(\beta I-\Lambda)^{-1} H\|_\infty = \max_i \left\{\frac{1}{|\beta-b_{ii}|} \sum_{j\neq i} |b_{ij}|\right\}.$$

This proves the Theorem. ■

Once again by the continuity argument used earlier, any connected component of the set consisting of the union of these disks contains as many eigenvalues of B as of Λ, i.e. if a connected component is formed out out k Gersgorin disks then it contains k eigenvalues of B.

Notes and references for Chapter 6

Theorem 23.1 was proved by Kahan in his paper [1]. Based on it was Theorem 23.3, which was proved by Kahan [2] among several other results. Up to dimension 16 the inequality (23.2) provides a better estimate than (23.4). For large n, of course the estimate (23.2) is much weaker. Kahan's estimates were improved by Schönage [1] who also refined the results in another direction deriving better inequalities using information about the sizes of the clusters of eigenvalues of A. (We have not touched upon such results in this monograph at all.) Theorem 23.2 was proved by Pokrzywa [1]. Example 23.4 occurs in the above papers of Kahan.

Theorem 25.1 was proved by Bauer and Fike [1]. Theorem 25.2 is known as the Gersgorin Disk Theorem and was proved by Gersgorin [1].

There are several perturbation results based on these two theorems. See the classical book of Wilkinson [1]. This book and the more recent book of Parlett [1] contain several results on perturbation bounds which involve some information about A, the separation of its eigenvalues from each other, its Jordan structure, etc. We also refer the reader to the papers of Jiang [1] and Kahan, Parlett and Jiang [1] for such results. Another important addition to recent literature is the book of Chatelin [1].

Postscripts

Between the writing of these notes and their going to press some further results on the topics discussed here have been announced/published. We give below a brief summary of these results arranged according to their relevance to the various sections of the text.

1. A postscript to Chapter 4, sections 13 and 14

In the preprint "Unitary invariance and spectral variation", R. Bhatia and J.A.R. Holbrook consider a class of norms wider than the unitarily-invariant norms. A norm τ from this class satisfies the weaker invariance property

$$\tau(UAU^*) = \tau(A)$$

for all unitary operators U. (See our remarks in Notes and references for Chapter 2).

In this paper Bhatia and Holbrook show that the "path-inequality" proved in Theorems 13.2 and 14.3 of Chapter 4 for the operator norm and for all unitarily-invariant norms, respectively, can be extended to this wider class of norms.

2. A postscript to Chapter 4, section 16

In the preprint "An extremal problem in Fourier analysis with applications to operator theory", R. Bhatia, C. Davis and P. Koosis show that the constant c_2 defined as in (16.8) of Chapter 4 is bounded as

$$\frac{\pi}{2} < c_2 < \frac{\pi}{2}\,\mathrm{Si}(\pi) < 2.91.$$

3. A postscript to Chapter 4, section 18

Let A be a Hermitian matrix and B a skew-Hermitian matrix with their respective eigenvalues $\alpha_1,\ldots,\alpha_n$ and $\beta_1,\ldots,\beta_n$ both arranged in decreasing order of modulus, $|\alpha_1| \geq \ldots \geq |\alpha_n|$ and $|\beta_1| \geq \ldots \geq |\beta_n|$. In the preprint "Eigenvalue inequalities associated with the Cartesian decomposition", T. Ando and R. Bhatia prove the following relations:

$$\left(\sum_{j=1}^{n} |\alpha_j - \beta_{n-j+1}|^p\right)^{1/p} \leq \|A-B\|_p \quad \text{for} \quad 2 \leq p \leq \infty,$$

$$\left(\sum_{j=1}^{n} |\alpha_j - \beta_j|^p\right)^{1/p} \leq 2^{1/p-1/2} \|A-B\|_p \quad \text{for} \quad 1 \leq p \leq 2,$$

where $\|\cdot\|_p$ denotes the Schatten p-norm for $1 \leq p \leq \infty$ with the convention $\|\cdot\|_\infty = \|\cdot\|$. In particular, these inequalities imply

$$\|(\text{Eig } A, \text{Eig } B)\|_p \leq \|A-B\|_p \quad \text{for} \quad 2 \leq p \leq \infty,$$

$$\|(\text{Eig } A, \text{Eig } B)\|_p \leq 2^{1/p-1/2} \|A-B\|_p \quad \text{for} \quad 1 \leq p \leq 2.$$

Notice that Theorem 18.1 of Chapter 4 due to Sunder is included in the first of these inequalities as a special case. The example given at the end of section 18 shows that none of the above inequalities can be improved.

It is reasonable to conjecture that for A, B as above we should have

$$|||(\text{Eig } A, \text{Eig } B)||| \leq \sqrt{2}\, |||A-B|||$$

For every unitarily-invariant norm. The authors of the paper mentioned above prove this when the eigenvalues of A and B lie in the first quadrant, i.e. when A and -iB are positive operators. In the general case, a weaker inequality with the constant 2 in place of $\sqrt{2}$ on the right hand side can be easily derived using the triangle inequality for norms.

4. A postscript to Chapter 5, sections 20 and 21

L. Elsner (1985) (see references section) has shown that the factor $n^{1/n}$ occurring on the right-hand sides of the inequalities (20.7) and (20.8) can be dropped. We outline Elsner's delightfully simple proof.

We will need to use the famous Hadamard Inequality which says that the absolute value of the determinant of a matrix is bounded by the product of the Euclidean norms of its column vectors.

Let A,B be two nxn matrices with eigenvalues $\alpha_1,\ldots,\alpha_n$ and $\beta_1,\ldots,\beta_n$, respectively. Define

$$\nu(\text{Spec } A, \text{ Spec } B) = \max_j \min_i |\alpha_i - \beta_j|$$

(See section 1).

Let, now, j denote the index at which the maximum in the above definition is attained. Choose an orthonormal basis $e_1,\ldots,e_n$ such that $Be_1 = \beta_j e_1$. Then

$$\begin{aligned}(\nu(\text{Spec } A, \text{ Spec } B))^n &= \min_i |\alpha_i - \beta_j|^n \\ &\leq \prod_{i=1}^{n} |\alpha_i - \beta_j| = |\det(A - \beta_j I)| \\ &\leq \|(A - \beta_j I)e_1\| \cdot \ldots \cdot \|(A - \beta_j I)\, e_n\|,\end{aligned}$$

by Hadamard's inequality. Now note that the first factor on the right-hand side of the above inequality can be written as $\|(A-B)e_1\|$ and is, therefore, bounded by $\|A-B\|$. The remaining n-1 factors can be bounded as $\|(A - \beta_j I)e_k\| \leq \|Ae_k\| + |\beta_j| \leq \|A\| + \|B\|$, for $k = 2,3,\ldots,n-1$. This shows

$$\nu(\text{Spec } A, \text{ Spec } B) \leq \|A-B\|^{1/n} (\|A\| + \|B\|)^{1-1/n}$$

Since the Hausdorff distance h(Spec A, Spec B) is defined as the maximum of ν(Spec A, Spec B) and ν(Spec B, Spec A) this shows

$$h(\text{Spec } A, \text{ Spec } B) \leq \|A-B\|^{1/n} (2M)^{1-1/n}$$

where $M = \max(\|A\|, \|B\|)$.

This is an improvement on the bound (20.7). The passage from this to an estimate for d(Eig A, Eig B) is effected by an argument like the one we have used in Chapter 5. This leads to an improvement of the bound (20.8) by knocking off the factor $n^{1/n}$ on the right-hand side.

References

Amir-Moez, A.R.

[1] Extreme properties of eigenvalues of a Hermitian transformation and singular values of the sum and product of linear transformations. Duke Math. J., 23(1956), 463-476.

[2] Extreme Properties of Linear Transformations and Geometry in Unitary Spaces, Texas Tech. University Math. Series Nos. 2 and 3, Lubbock, Texas, 1968.

Ando, T.

[1] Majorization, Doubly Stochastic Matrices and Comparison of Eigenvalues, Mimeographed Lecture Notes, Sapporo, Japan, 1982.

Arnol'd, V.I.

[1] On matrices depending on parameters, Uspekhi Mat. Nauk, 26 (2) (1971); Russian Math. Surveys, 26 (2) (1971), 29-43.

Auslander, L. and MacKenzie, R.

[1] Introduction to Differentiable Manifolds, McGraw-Hill, New York,1963,

Bauer, F.L. and Fike, C.T.

[1] Norms and exclusion theorems, Numer. Math., 2(1960), 137-141.

Baumgärtel, H.

[1] Endlichdimensionale Analytische Störungstheorie, Akademie-Verlag, Berlin, 1972.

Beckenback, E.F. and Bellman, R.

[1] Inequalities, Springer, Berlin, 1971.

Bellman, R.

[1] Introduction to Matrix Analysis, McGraw-Hill, New York, 1960.

Berezin, F.A. and Gel'fand, I.M.

[1] Some remarks on the theory of spherical functions on symmetric Riemannian manifolds, Trudy Moskov. Mat. Obsc., 5(1956), 311-351.

Bhatia, R.

[1] Analysis of spectral variation and some inequalities, Trans. Amer. Math. Soc., 272(1982), 323-332.

[2] On the rate of change of spectra of operators II, Linear Algebra and Appl.,36(1981), 25-32.

[3] Variation of symmetric tensor powers and permanents, Linear Algebra and Appl., 62 (1984), 269-276.

Bhatia, R. and Davis, C.

[1] A bound for the spectral variation of a unitary operator, Linear and Multilinear Algebra, 15 (1984), 71-76.

Bhatia, R., Davis, C. and McIntosh, A.

[1] Perturbation of spectral subspaces and solution of linear operator equations, Linear Algebra and Appl., 52-53 (1983), 45-67.

Bhatia, R. and Friedland, S.

[1] Variation of Grassman powers and spectra, Linear Algebra and Appl., 40 (1981), 1-18.

Bhatia, R. and Holbrook, J.A.R.

[1] Short normal paths and spectral variation, Proc. Amer. Math. Soc., 94 (1985), 377-382.

Bhatia, R. and Mukherjea, K.K.

[1] On the rate of change of spectra of operators, Linear Algebra and Appl., 27 (1979), 147-157.

[2] The space of unordered tuples of complex numbers, Linear Algebra and Appl., 52 (1983), 765-768.

Bhatia, R. and Parthasarathy, K.R.

[1] Lectures on Functional Analysis, McMillan Co., New Delhi, 1978.

Birkhoff, G.D.

[1] Tres observaciones sobre el algebra lineal, Univ. Nac. Tucuman Rev. Ser. A., 5 (1946), 147-151.

Blokhuis, A. and Seidel, J.J.

[1] An introduction to multilinear algebra and some applications, Philips J. Research, 39 (1984), 111-120.

Chatelin, F.

[1] Spectral Approximation of Linear Operators, Academic Press, New York, 1983.

Collatz, L.

[1] Functional Analysis and Numerical Mathematics, Academic Press, New York, 1966.

Courant, R.

[1] Über die Eigenwerte bei den Differentialgleichungen der mathematischen Physik, Math. Z., 7 (1920), 1-57.

Courant, R. and Hilbert, D.

[1] Methods of Mathematical Physics, first published in German by Springer, Berlin, 1924 and in an English translation by Wiley, New York, 1953.

Davis, C.

[1] Various averaging operations onto subalgebras, Illinois J. Math., 3 (1959), 538-553.

[2] The rotation of eigenvectors by a perturbation, J. Math. Anal. Appl., 6 (1963), 159-173.

[3] The rotation of eigenvectors by a perturbation II, J. Math. Anal. Appl., 11 (1965), 20-27.

Davis, C. and Kahan, W.M.

[1] The rotation of eigenvectors by a perturbation III, SIAM J. Numer. Anal., 7 (1968), 1-46.

Dieudonné, J.

[1] Foundations of Modern Analysis, Academic Press, New York, 1960.

Elsner, L.

[1] On the variation of the spectra of matrices, Linear Algebra and Appl., 47 (1982), 127-138.

[2] An optimal bound for the spectral variation of two matrices, Linear Algebra and Appl., 71 (1985), 77-80.

Elsner, L., Johnson, C., Ross, J. and Schönheim, J.

[1] On a generalised matching problem arising in estimating the eigenvalue variation of two matrices, Europ. J. Combinatorics, 4 (1983), 133-136.

Fan, K.

[1] On a theorem of Weyl concerning eigenvalues of linear transformations I, Proc. Nat. Acad. Sci. U.S.A., 35 (1949), 652-655.

[2] On a theorem of Weyl concerning eigenvalues of linear transformations II, Proc. Nat. Acad. Sci. U.S.A., 36 (1950), 31-35.

[3] Maximum properties and inequalities for the eigenvalues of completely continuous operators, Proc. Nat. Acad. Sci. U.S.A., 37 (1951), 760-766.

Fan, K. and Hoffman, A.J.

[1] Some metric inequalities in the space of matrices, Proc. Amer. Math. Soc., 6 (1955), 111-116.

Fischer, E.

[1] Über quadratische Formen mit reelen Koeffizienten, Monatsh. Math. Phys., 16 (1905), 234-248.

Fong, C.K. and Holbrook, J.A.R.

[1] Unitarily invariant operator norms, Canadian J. Math., 35 (1983), 274-299.

Friedland, S.

[1] Variation of tensor powers and spectra, Linear and Multilinear Algebra, 12 (1982), 81-98.

Gårding, L.

[1] Linear hyperbolic partial differential equations with constant coefficients, Acta Math., 85 (1951), 1-62.

[2] An inequality for hyperbolic polynomials, J. Math. and Mech., 8 (1959), 957-966.

Gersgorin, S.A.

[1] Über die Abgrenzung der Eigenwerte einer Matrix, Izv. Akad. Nauk. SSSR, Ser. Fiz. - Mat., 6 (1931), 749-754.

Gerstenhaber, M.

[1] Note on a theorem of Wielandt, Math. Z., 71 (1959), 141-142.

Gohberg, I.C. and Krein, M.G.

[1] Introduction to the Theory of Linear Nonselfadjoint operators, Transl. Math. Monographs, Vol. 18, Amer. Math. Soc., Providence, R.I., 1969.

Greub, W.

[1] Multilinear Algebra, Springer, Berlin, 1967.

Hall, P.

[1] On representatives of subsets, J. London Math. Soc., 10 (1935), 26-30.

Halmos, P.R.

[1] Finite-dimensional Vector Spaces, Van Nostrand, Princeton, 1958.

[2] A Hilbert Space Problem Book, Springer, Berlin, 1974.

Hardy, G.H., Littlewood, J.E. and Pólya, G.

[1] Inequalities, Cambridge Univ. Press, 1934.

Henrici, P.

[1] Bounds for iterates, inverses, spectral variation and fields of values of nonnormal matrices, Numer. Math., 4 (1962), 24-39.

Hoffman, A.J. and Wielandt, H.W.

[1] The variation of the spectrum of a normal matrix, Duke Math. J., 20 (1953), 37-39.

Householder, A.S.

[1] The Theory of Matrices in Numerical Analysis, Ginn (Blaisdell), Boston, 1964.

Jiang, E.

[1] On spectral variation of a nonnormal matrix, Linear Algebra and Applications, 42 (1983), 223-241.

Kahan, W.

[1] Every nxn matrix Z with real spectrum satisfies $\|Z-Z^*\| \leq \|Z+Z^*\| (\log_2 n + 0.038)$, Proc. Amer. Math. Soc., 39 (1973), 235-241.

[2] Spectra of nearly Hermitian matrices, Proc. Amer. Math. Soc., 48 (1975), 11-17.

Kahan, W., Parlett, B.N. and Jiang, E.

[1] Residual bounds on approximate eigensystems of nonnormal matrices, SIAM J. Numer. Anal., 19 (1981), 470-484.

Kato, T.

[1] A Short Introduction to Perturbation Theory for Linear Operators, Springer, Berlin, 1982.

Lang, S.

[1] Algebra, Addison-Wesley, Reading, 1970.

Lax, P.D.

[1] Differential equations, difference equations and matrix theory, Comm. Pure and Appl. Math., 11 (1958), 175-194.

Lidskii, V.B.

[1] The proper values of the sum and product of symmetric matrices, Dokl. Akad. Nauk SSSR, 75 (1950), 769-772.

Marcus, M.

[1] Finite-dimensional Multilinear Algebra, Marcel Dekker, New York, 1973.

Marcus, M. and Minc, H.

[1] A Survey of Matrix Theory and Matrix Inequalities, Prindle, Weber and Schmidt, Boston, 1964.

Marshall, A.W. and Olkin, I.

[1] Inequalities : Theory of Majorisation and its Applications, Academic Press, New York, 1979.

Mirsky, L.

[1] Symmetric gauge functions and unitarily invariant norms, Quart. J. Math. Oxford Ser. (2), 11 (1960), 50-59.

[2] Inequalities for normal and Hermitian matrices, Duke Math. J., 24 (1957), 591-599.

Ostrowski, A.

[1] Recherches sur la méthode de Gräffe et les zeros des polynômes et des series de Laurent, Acta Math., 72 (1940), 99-257.

[2] Über die Stetigkeit von charakteristischen Wurzeln in Abhängigkeit von den Matrizenelementen, Jber. Deut. Mat. - Verein, 60 (1957), 40-42.

[3] Solution of Equations and Systems of Equations, Academic Press, New York 1960.

Parlett, B.N.

[1] The Symmetric Eigenvalue Problem, Prentice-Hall, Englewood Cliffs, New Jersey, 1980.

Parthasarathy, K.R.

[1] Eigenvalues of matrix-valued analytic maps, J. Austral. Math. Soc. (Ser. A), 26 (1978), 179-197.

Pokrzywa, A.

[1] Spectra of operators with fixed imaginary parts, Proc. Amer. Math. Soc., 81 (1981), 359-364.

Poston, T. and Stewart, I.

[1] Catastrophe Theory and its Applications, Pitman, London, 1978.

Radjavi, H. and Rosenthal, P.

[1] Invariant Subspaces, Springer, Berlin, 1973.

Rao, C.R.

[1] Linear Statistical Inference, Wiley, New York, 1973.

Rellich, F.

[1] Perturbation Theory of Eigenvalue Problems, Gordon and Breach, New York, 1969.

Riddell, R.C.

[1] Minimax Problems on Grassmann Manifolds. Sums of Eigenvalues, Advances in Math., 54 (1984), 107-199.

Rosenblum, M.

[1] On the operator equation BX-XA = Q, Duke Math. J., 23 (1956), 263-269.

Schatten, R.

[1] Norm Ideals of Completely Continuous Operators, Springer, Berlin, 1960.

Schönhage, A.

[1] Arbitrary perturbations of Hermitian matrices, Linear Algebra and Applications, 24 (1979), 143-149.

Schur, I.

[1] Über eine Klasse von Mittelbildungen mit Anwendungen auf die Determinanten Theorie, Sitz. Berlin Math. Gesellschaft, 22 (1923), 9-20.

Simon, B.

[1] Trace Ideals and their Applications, London Math. Soc. Lecture Note Series; 35, Cambridge Univ. Press, 1979.

Stewart, G.W.

[1] Introduction to Matrix Computations, Academic Press, New York, 1973.

Sunder, V.S.

[1] On permutations, convex hulls and normal operators, Linear Algebra and Appl., 48 (1982), 403-411.

[2] Distance between normal operators, Proc. Amer. Math. Soc., 84 (1982), 483-484.

Thompson, R.C. and Freede, L.

[1] On the eigenvalues of the sum of Hermitian matrices, Linear Algebra and Appl., 4 (1971), 369-376.

von Neumann, J.

[1] Some matrix inequalities and metrization of matric space, Tomsk. Univ. Rev., 1 (1937), 286-300.

Weinberger, H.F.

[1] Remarks on the preceding paper of Lax, Comm. Pure and Appl. Math., 11 (1958), 195-196.

Weyl, H.

[1] Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen, Math. Ann. 71 (1912), 441-479.

[2] Inequalities between the two kinds of eigenvalues of a linear transformation,Proc. Nat.Acad.Sci. U.S.A., 35 (1949), 408-411.

Wielandt, H.W.

[1] Lineare Scharen von Matrizen mit reellen Eigenwerten, Math. Z., 53 (1950), 219-225.

[2] An extremum property of sums of eigenvalues, Proc. Amer. Math. Soc., 6 (1955), 106-110.

Wilkinson, J.H.

[1] The Algebraic Eigenvalue Problem, Oxford Univ. Press, 1965.

[2] The evaluation of the zeros of ill-conditioned polynomials, Numerische Math., 1 (1959), 150-166.

Wilson, R.J.

[1] Introduction to Graph Theory, Oliver and Boyd, London, 1972.

Supplements 1986-2006

7 Singular values and norms

§26 The minmax principle

One of the important corollaries of the minmax principle is Cauchy's Interlacing Theorem. This theorem gives interesting inequalities between the eigenvalues of a Hermitian matrix A and those of a principal submatrix B.

Let A be a Hermitian operator on an n-dimensional Hilbert space $\mathcal{H}$. Let $\mathcal{N}$ be an $(n-k)$-dimensional subspace of $\mathcal{H}$, and let V be the injection map from $\mathcal{N}$ into $\mathcal{H}$. Then $B = V^*AV$ is a linear operator on $\mathcal{N}$, and is called the *compression* of A to the subspace $\mathcal{N}$. In an appropriate coordinate system A has a matrix representation

$$A = \begin{bmatrix} B & * \\ * & * \end{bmatrix}$$

in which B is the $(n-k) \times (n-k)$ block in the top left corner of the $n \times n$ matrix A.

Theorem 26.1 (Cauchy's Interlacing Theorem) Let A be a Hermitian operator on $\mathcal{H}$ and let B be its compression to an $(n-k)$-dimensional subspace $\mathcal{N}$. Then for $j = 1, 2, \ldots, n-k$

$$\lambda_{[j]}(A) \geq \lambda_{[j]}(B) \geq \lambda_{[j+k]}(A). \tag{26.1}$$

Proof Let $1 \leq j \leq n-k$, and let $\mathcal{M}$ be the j-dimensional space spanned by the eigenvectors of B corresponding to its eigenvalues $\lambda_{[1]}(B), \ldots, \lambda_{[j]}(B)$. Then $\langle Bx, x\rangle = \langle Ax, x\rangle$ for all $x \in \mathcal{M}$, and hence

$$\lambda_{[j]}(B) = \min_{x\in\mathcal{M}, \|x\|=1} \langle Bx, x\rangle = \min_{x\in\mathcal{M}, \|x\|=1} \langle Ax, x\rangle.$$

So, by the minmax principle (Theorem 6.1)

$$\lambda_{[j]}(B) \leq \lambda_{[j]}(A), \quad 1 \leq j \leq n-k.$$

This is the first inequality in (26.1). Replace in this inequality A and B by their negatives. Then observe that

$$\lambda_{[j]}(-A) = -\lambda_{[n-j+1]}(A) \quad \text{for} \quad 1 \leq j \leq n,$$

and

$$\lambda_{[j]}(-B) = -\lambda_{[n-k-j+1]}(B) \quad \text{for} \quad 1 \leq j \leq n-k.$$

This leads to the inequality

$$\lambda_{[n-k-j+1]}(B) \geq \lambda_{[n-j+1]}(A) \quad \text{for} \quad 1 \leq j \leq n-k,$$

which on renaming indices becomes

$$\lambda_{[i]}(B) \geq \lambda_{[i+k]}(A) \quad \text{for} \quad 1 \leq i \leq n-k.$$

This is the second inequality in (26.1). ■

These inequalities have a particularly attractive form when B is the compression of A to an $(n-1)$ dimensional subspace. In this case if the eigenvalues of A are enumerated as $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_n$, and those of B as $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_{n-1}$, then the interlacing theorem says

$$\alpha_1 \geq \beta_1 \geq \alpha_2 \geq \cdots \geq \beta_{n-1} \geq \alpha_n. \tag{26.2}$$

This "interlacing" of the two sets of eigenvalues gives the theorem its name.

Another important and useful theorem that can be easily derived from the minmax principle is Sylvester's law of inertia.

The *inertia* of an $n \times n$ Hermitian matrix A is the triple

$$\operatorname{In}(A) = (\pi(A), \zeta(A), \nu(A)),$$

where $\pi(A)$, $\zeta(A)$, $\nu(A)$ are nonnegative integers that count the positive, zero, and negative eigenvalues of A (counted with multiplicities). Two $n \times n$ matrices A and B are said to be *congruent* if there exists an invertible matrix X such that $B = X^*AX$.

Theorem 26.2 (Sylvester's Law of Inertia) Two Hermitian matrices A and B are congruent if and only if $\operatorname{In}(A) = \operatorname{In}(B)$.

Proof By the minmax principle $\lambda_{[j]}(A) > 0$ if and only if there exists a j-dimensional subspace $\mathcal{M}$ such that

$$\min_{u \in \mathcal{M}, u \neq 0} \langle Au, u \rangle > 0.$$

If X is an invertible operator, then the space $\mathcal{N} = X^{-1}(\mathcal{M})$ has dimension j, and

$$\min_{u \in \mathcal{M}, u \neq 0} \langle Au, u \rangle = \min_{v \in \mathcal{N}, v \neq 0} \langle AXv, Xv \rangle = \min_{v \in \mathcal{N}, v \neq 0} \langle X^* AXv, v \rangle.$$

This argument shows that if $B = X^*AX$, then $\pi(A) = \pi(B)$. By the same argument $\nu(A) = \nu(B)$, and hence $\mathrm{In}\,(A) = \mathrm{In}\,(B)$.

It is easy to see that any Hermitian matrix with inertia (π, ζ, ν) is congruent to the diagonal matrix with entries 1, 0, -1 occurring π, ζ, and ν times on its diagonal. Thus two Hermitian matrices with equal inertias are congruent. ■

Let A be a Hermitian matrix partitioned as

$$A = \begin{bmatrix} H_1 & E \\ E^* & H_2 \end{bmatrix}, \tag{26.3}$$

where H_1 and H_2 are Hermitian matrices, possibly of different sizes. Suppose H_2 is invertible, then

$$\begin{bmatrix} I & -EH_2^{-1} \\ O & I \end{bmatrix} \begin{bmatrix} H_1 & E \\ E^* & H_2 \end{bmatrix} \begin{bmatrix} I & O \\ -H_2^{-1}E^* & I \end{bmatrix} = \begin{bmatrix} H_1 - EH_2^{-1}E^* & O \\ O & H_2 \end{bmatrix}. \tag{26.4}$$

The matrix $\widetilde{H}_2 = H_1 - EH_2^{-1}E^*$ is called the *Schur complement* of H_2 in A. The matrix A is congruent to the block diagonal matrix

$$\begin{bmatrix} \widetilde{H}_2 & O \\ O & H_2 \end{bmatrix}.$$

In particular, this shows that

$$\mathrm{In}\,(A) = \mathrm{In}\,(\widetilde{H}_2) + \mathrm{In}\,(H_2). \tag{26.5}$$

§27 Symmetric gauge functions and norms

To the discussion in Section 7 we add some techniques that have been found to be especially useful in proving inequalities involving unitarily invariant norms.

Let Φ be any norm on $\mathbb{R}^n$. Its *dual* defined by the relation

$$\Phi'(x) = \max_{\Phi(y)=1} |\langle x, y\rangle| = \max_{\Phi(y)=1} \left|\sum_{i=1}^{n} x_i y_i\right|$$

is another norm on $\mathbb{R}^n$. Obviously $\Phi'' = \Phi$, and it is easy to see that if a norm is a symmetric gauge function, then so is its dual norm.

Let Φ be a symmetric gauge function, and let $|||\cdot|||_\Phi$ be the unitarily-invariant norm it induces on $\mathbb{M}(n)$. Then we have

$$|||A|||_\Phi = \Phi(s_1(A), \ldots, s_n(A)) = \max_{\Phi'(\alpha)=1} \left|\sum_{i=1}^{n} \alpha_i s_i(A)\right|.$$

It is not difficult to see that

$$|||A|||_\Phi = \max_{\alpha \in \mathbb{R}^n_{+\downarrow}, \Phi'(\alpha)=1} \sum_{i=1}^{n} \alpha_i s_i(A), \tag{27.1}$$

where $\mathbb{R}^n_{+\downarrow}$ is the set of all n-vectors whose coordinates are decreasingly ordered nonnegative numbers.

For each $\alpha \in \mathbb{R}^n_{+\downarrow}$ let

$$\|A\|_\alpha = \sum_{i=1}^{n} \alpha_i s_i(A). \tag{27.2}$$

Then $\alpha_1^{-1}\|A\|_\alpha$ is a unitarily-invariant norm (the factor α_1^{-1} has the effect of normalisation in keeping with our convention). The relation (27.1) shows that every unitarily-invariant norm has a representation

$$|||A|||_\Phi = \max_{\alpha \in K_\Phi} \|A\|_\alpha, \tag{27.3}$$

where K_Φ is the compact subset of $\mathbb{R}^n_{+\downarrow}$ consisting of all α for which $\Phi'(\alpha) = 1$.

Let $\|A\|_k$, $1 \le k \le n$, be the family of Ky Fan norms. Then we have the identity

$$\|A\|_\alpha = \sum_{k=1}^{n} (\alpha_k - \alpha_{k+1}) \|A\|_k, \tag{27.4}$$

for all $\alpha \in \mathbb{R}^n_{+\downarrow}$, with the convention that $\alpha_{n+1} = 0$.

So, if A and B are two operators such that $\|A\|_k \le \|B\|_k$ for $1 \le k \le n$, then (27.4) shows that $\|A\|_\alpha \le \|B\|_\alpha$ for all $\alpha \in \mathbb{R}^n_{+\downarrow}$, and it follows from (27.3) that $|||A||| \le |||B|||$ for every unitarily-invariant norm. Thus we have a simple alternative proof for Theorem 7.5.

Another illustration of the use of the norms $\|\cdot\|_\alpha$ is given in the following proof.

Proposition 27.1 Let $A, B,$ and C be operators such that $\|A\|_k^2 \le \|B\|_k \|C\|_k$ for all Ky Fan norms. Then $|||A|||^2 \le |||B|||\,|||C|||$ for all unitarily-invariant norms.

Proof The given condition may be stated in another way: for each k, the 2×2 matrix

$$\begin{bmatrix} \|B\|_k & \|A\|_k \\ \|A\|_k & \|C\|_k \end{bmatrix}$$

is positive semidefinite. This implies that for each $\alpha \in \mathbb{R}^n_{+\downarrow}$, the matrix

$$\begin{bmatrix} \|B\|_\alpha & \|A\|_\alpha \\ \|A\|_\alpha & \|C\|_\alpha \end{bmatrix} = \sum_{k=1}^{n} (\alpha_k - \alpha_{k+1}) \begin{bmatrix} \|B\|_k & \|A\|_k \\ \|A\|_k & \|C\|_k \end{bmatrix}$$

is positive semidefinite. This shows that $\|A\|_\alpha^2 \le \|B\|_\alpha \|C\|_\alpha$ for all α, and hence by (27.3) $|||A|||^2 \le |||B|||\,|||C|||$ for every unitarily-invariant norm. ■

The next theorem gives yet another representation of Ky Fan norms.

Theorem 27.2 For $k = 1, 2, \ldots, n$ we have

$$\|A\|_k = \min \{\|B\|_{\text{tr}} + \|C\| : B + C = A\}. \tag{27.5}$$

Proof By definition

$$\|X\| = s_1(X) \le \sum_{j=1}^{k} s_j(X) \le \sum_{j=1}^{n} s_j(X) = \|X\|_{\text{tr}}.$$

§27 Symmetric gauge functions and norms

To the discussion in Section 7 we add some techniques that have been found to be especially useful in proving inequalities involving unitarily invariant norms.

Let Φ be any norm on $\mathbb{R}^n$. Its *dual* defined by the relation

$$\Phi'(x) = \max_{\Phi(y)=1} |\langle x, y\rangle| = \max_{\Phi(y)=1} \left|\sum_{i=1}^{n} x_i y_i\right|$$

is another norm on $\mathbb{R}^n$. Obviously $\Phi'' = \Phi$, and it is easy to see that if a norm is a symmetric gauge function, then so is its dual norm.

Let Φ be a symmetric gauge function, and let $\|\cdot\|_\Phi$ be the unitarily-invariant norm it induces on $\mathbb{M}(n)$. Then we have

$$|||A|||_\Phi = \Phi(s_1(A), \ldots, s_n(A)) = \max_{\Phi'(\alpha)=1} \left|\sum_{i=1}^{n} \alpha_i s_i(A)\right|.$$

It is not difficult to see that

$$|||A|||_\Phi = \max_{\alpha\in\mathbb{R}^n_{+\downarrow}, \Phi'(\alpha)=1} \sum_{i=1}^{n} \alpha_i s_i(A), \tag{27.1}$$

where $\mathbb{R}^n_{+\downarrow}$ is the set of all n -vectors whose coordinates are decreasingly ordered nonnegative numbers.

For each $\alpha \in \mathbb{R}^n_{+\downarrow}$ let

$$\|A\|_\alpha = \sum_{i=1}^{n} \alpha_i s_i(A). \tag{27.2}$$

Then $\alpha_1^{-1}\|A\|_\alpha$ is a unitarily-invariant norm (the factor α_1^{-1} has the effect of normalisation in keeping with our convention). The relation (27.1) shows that every unitarily-invariant norm has a representation

$$|||A|||_\Phi = \max_{\alpha\in K_\Phi} \|A\|_\alpha, \tag{27.3}$$

where K_Φ is the compact subset of $\mathbb{R}^n_{+\downarrow}$ consisting of all α for which $\Phi'(\alpha) = 1$.

Let $\|A\|_k$, $1 \le k \le n$, be the family of Ky Fan norms. Then we have the identity

$$\|A\|_\alpha = \sum_{k=1}^{n} (\alpha_k - \alpha_{k+1}) \|A\|_k, \tag{27.4}$$

for all $\alpha \in \mathbb{R}^n_{+\downarrow}$, with the convention that $\alpha_{n+1} = 0$.

So, if A and B are two operators such that $\|A\|_k \le \|B\|_k$ for $1 \le k \le n$, then (27.4) shows that $\|A\|_\alpha \le \|B\|_\alpha$ for all $\alpha \in \mathbb{R}^n_{+\downarrow}$, and it follows from (27.3) that $|||A||| \le |||B|||$ for every unitarily-invariant norm. Thus we have a simple alternative proof for Theorem 7.5.

Another illustration of the use of the norms $\|\cdot\|_\alpha$ is given in the following proof.

Proposition 27.1 Let A, B, and C be operators such that $\|A\|_k^2 \le \|B\|_k \|C\|_k$ for all Ky Fan norms. Then $|||A|||^2 \le |||B|||\,|||C|||$ for all unitarily-invariant norms.

Proof The given condition may be stated in another way: for each k, the 2×2 matrix

$$\begin{bmatrix} \|B\|_k & \|A\|_k \\ \|A\|_k & \|C\|_k \end{bmatrix}$$

is positive semidefinite. This implies that for each $\alpha \in \mathbb{R}^n_{+\downarrow}$, the matrix

$$\begin{bmatrix} \|B\|_\alpha & \|A\|_\alpha \\ \|A\|_\alpha & \|C\|_\alpha \end{bmatrix} = \sum_{k=1}^{n} (\alpha_k - \alpha_{k+1}) \begin{bmatrix} \|B\|_k & \|A\|_k \\ \|A\|_k & \|C\|_k \end{bmatrix}$$

is positive semidefinite. This shows that $\|A\|_\alpha^2 \le \|B\|_\alpha \|C\|_\alpha$ for all α, and hence by (27.3) $|||A|||^2 \le |||B|||\ |||C|||$ for every unitarily-invariant norm. ■

The next theorem gives yet another representation of Ky Fan norms.

Theorem 27.2 For $k = 1, 2, \ldots, n$ we have

$$\|A\|_k = \min \{\|B\|_{\text{tr}} + \|C\| : B + C = A\}. \tag{27.5}$$

Proof By definition

$$\|X\| = s_1(X) \le \sum_{j=1}^{k} s_j(X) \le \sum_{j=1}^{n} s_j(X) = \|X\|_{\text{tr}}.$$

So, if B and C are any two operators such that $A = B + C$, then

$$\|A\|_k \leq \|B\|_k + \|C\|_k \leq \|B\|_{\text{tr}} + k\|C\|.$$

Let A have the singular value decomposition $A = USV$, where $S = \text{diag}(s_1, \ldots, s_n)$. Put $B = U\Gamma V$, and $C = U\Lambda V$ with

$$\Gamma = \text{diag}(s_1 - s_k, s_2 - s_k, \ldots, s_k - s_k, 0, \ldots, 0),$$

and

$$\Lambda = \text{diag}(s_k, s_k, \ldots, s_k, s_{k+1}, s_{k+2}, \ldots, s_n),$$

then $B + C = A$. In this case $\|B\|_{\text{tr}} = \sum_{j=1}^{k} s_j - ks_k = \|A\|_k - ks_k$, and $\|C\| = s_k$. Thus $\|A\|_k = \|B\|_{\text{tr}} + k\|C\|$ for this particular choice of B and C. ■

Let us denote by $\{s_j(A)\}$ the n-tuple of singular values of A. We have seen that the weak majorisation

$$\{s_j(A)\} \prec_w \{s_j(B)\} \tag{27.6}$$

is equivalent to the condition

$$|||A||| \leq |||B||| \quad \text{for all unitarily-invariant norms.} \tag{27.7}$$

Since $f(x) = x^2$ is convex and monotone increasing on $\mathbb{R}_+$, the condition (27.6) implies that

$$\{s_j^2(A)\} \prec_w \{s_j^2(B)\}. \tag{27.8}$$

In several cases the weak majorisation (27.8) is true but not (27.6). It is useful to translate this to norm inequalities.

We say that a unitarily-invariant norm $|||\cdot|||$ is a Q*-norm* if there exists another unitarily-invariant norm $|||\cdot|||^{\wedge}$ such that

$$|||A|||^2 = |||A^*A|||^{\wedge} \quad \text{for all } A. \tag{27.9}$$

A Schatten p-norm is a Q-norm if and only if $p \geq 2$, because

$$\|A\|_p^2 = \|A^*A\|_{p/2}.$$

The condition (27.8) is equivalent to the following inequalities for norms

$$\|A\|_Q \leq \|B\|_Q \quad \text{for all } Q\text{-norms.} \tag{27.10}$$

Notes and references

To the list of basic books on numerical linear algebra we should add J. Demmel [D], G. H. Golub and C. F. Van Loan [GV], and N. J. Higham [Hi]. Closer to our book in content and spirit is the work G.W. Stewart and J.-G. Sun[SS]. Closer still are Chapters 6-8 of [B1].

The norms (27.2) seem to have been introduced as "generalized spectral norms" by C.-K. Li, T.-Y. Tam and N.-K. Tsing [LTS]. They were used to good effect by R. A. Horn and R. Mathias [HM] in proving inequalities for unitarily-invariant norms. The ideas of this paper recur in C.-K. Li and R. Mathias [LM1] from whom we have taken the proof of Proposition 27.1. An alternate proof due to T. Ando and F. Hiai is given in R. Bhatia, F. Kittaneh and R.-C. Li [BKL] where this problem arose in the first place. A good use of the representation (27.5) was made by F. Hiai and Y. Nakamura [HN] after whom other authors used it. The idea that norms like (27.9) are special occurs in C. Davis and W. M. Kahan[DK], page 22. In the mid 1980's there was vigorous activity around majorization and inequalities for unitarily-invariant norms. Looking for majorization principles behind some famous inequalities for the Schatten p-norms (like Clarkson's) that change direction as the condition $p \geq 2$ is replaced by $p \leq 2$, this author introduced the terminology Q-norms in [B2]. After that it was observed in several papers that inequalities that are true for all p-norms are often true for all unitarily-invariant norms, those that are valid only under the restrictions $p \geq 2$, or $p \leq 2$ are often true for all Q-norms and their dual norms, respectively. See, for example, [AB], [BH2] and [BH3].

8 Spectral variation of Hermitian matrices

§28 Bounds in the Frobenius norm

In 1934 Karl Löwner (later Charles Loewner) wrote a most remarkable paper that is widely known for initiating the theory of matrix monotone functions. In this paper Löwner states, without proof, the Frobenius norm analogue of Theorem 8.5. As we have seen, this was subsequently generalized in two different directions: Theorem 9.7 for Hermitian matrices asserting the same inequality for all unitarily-invariant norms, and Theorem 15.1 of Hoffman and Wielandt valid for normal matrices but restricted to the Frobenius norm. We present two proofs of Löwner's theorem that depend on ideas simpler than the ones needed for these more general versions.

Lemma 28.1 Let x and y be any two vectors in $\mathbb{R}^n$. Then

$$\langle x_\downarrow, y_\uparrow \rangle \leq \langle x, y \rangle \leq \langle x_\downarrow, y_\downarrow \rangle. \tag{28.1}$$

Proof It is enough to prove this for $n = 2$. In this case the assertion is that whenever $x_1 \geq x_2$, and $y_1 \geq y_2$, then $x_1 y_1 + x_2 y_2 \geq x_1 y_2 + x_2 y_1$. The latter inequality can be written as $(x_1 - x_2)(y_1 - y_2) \geq 0$ and is obviously true. ■

A matrix version of this is the following

Proposition 28.2 Let A and B be $n \times n$ Hermitian matrices. Then

$$\langle \mathrm{Eig}_\downarrow(A), \mathrm{Eig}_\uparrow(B) \rangle \;\leq\; \mathrm{tr}\, AB \;\leq\; \langle \mathrm{Eig}_\downarrow(A), \mathrm{Eig}_\downarrow(B) \rangle. \tag{28.2}$$

Proof If A and B were commuting matrices, this would reduce to the preceding Lemma. The general case, in turn, can be reduced to this special one as follows.

Let $U(n)$ be the group consisting of $n \times n$ unitary matrices and let

$$\mathcal{U}_B = \{UBU^* : U \in U(n)\}.$$

If we replace B by any element of $\mathcal{U}_B$, then $\operatorname{Eig}(B)$ is not changed, and hence nor are the two inner products in (28.2). Consider the function

$$f(X) = \operatorname{tr} AX, \quad X \in \mathcal{U}_B.$$

The two inequalities in (28.2) are, in fact, lower and upper bounds for $f(X)$. We prove them by showing that the maximum and the minimum of f are attained at matrices that commute with A. In fact we will prove more than this: if X_0 is any extreme point for f, then X_0 commutes with A.

If a point X_0 on $\mathcal{U}_B$ is an extreme point for f, then

$$\left.\frac{d}{dt}\right|_{t=0} \operatorname{tr} AU(t)X_0U(t)^* = 0$$

for every differentiable curve $U(t)$ with $U(0) = I$. Equivalently,

$$\left.\frac{d}{dt}\right|_{t=0} \operatorname{tr} Ae^{tK}X_0e^{-tK} = 0$$

for every skew-Hermitian matrix K. Expanding the exponential function into a series, we see that this condition reduces to

$$\operatorname{tr}\,(AKX_0 - AX_0K) = 0.$$

By the cyclicity of the trace this is the same as the statement

$$\operatorname{tr} K\,(X_0A - AX_0) = 0. \tag{28.3}$$

On the space of skew-Hermitian matrices $\langle K, L\rangle = -\operatorname{tr} KL$ is an inner product. So, if (28.3) is valid for all skew-Hermitian K, then we must have $X_0A - AX_0 = 0$. ■

Second Proof We can apply a unitary similarity and assume that A is diagonal, and further $A = \operatorname{Eig}_\downarrow(A) = \operatorname{diag}(\alpha_1, \ldots, \alpha_n)$. Then

$$\operatorname{tr} AB = \sum_{i=1}^{n} \alpha_i d_i = \langle \alpha, d\rangle,$$

where $d = (d_1, \dots, d_n)$ is the diagonal of B.

By Corollary 6.8 the vector d is majorised by $\lambda(B)$, the vector whose coordinates are the eigenvalues of B. It follows from Theorem 3.1 that d lies in the convex hull of the vectors $\lambda_\sigma(B)$ whose coordinates are permutations of the coordinates of $\lambda(B)$. On this convex set Ω, the function $f(\omega) = \sum_{i=1}^{n} \alpha_i \omega_i$ is affine, and hence attains its maximum and minimum on the vertices of Ω. These vertices are among the points $\lambda_\sigma(B)$. So the Proposition follows from Lemma 28.1. ■

Theorem 28.3 Let A and B be Hermitian matrices. Then

$$\|\mathrm{Eig}_\downarrow(A) - \mathrm{Eig}_\downarrow(B)\|_F \;\le\; \|A - B\|_F \;\le\; \|\mathrm{Eig}_\downarrow(A) - \mathrm{Eig}_\uparrow(B)\|_F. \quad (28.4)$$

Proof Let $\alpha_1, \dots, \alpha_n$, and $\beta_1, \dots, \beta_n$ be the eigenvalues of A and B, respectively. Then

$$\begin{aligned} \|A - B\|_F^2 &= \|A\|_F^2 + \|B\|_F^2 - 2\,\mathrm{tr}\,AB \\ &= \sum_{i=1}^{n} \alpha_i^2 + \sum_{i=1}^{n} \beta_i^2 - 2\,\mathrm{tr}\,AB \end{aligned}$$

By the second inequality in Proposition 28.2

$$\|A - B\|_F^2 \;\ge\; \sum_{i=1}^{n} \alpha_i^2 + \sum_{i=1}^{n} \beta_i^2 - 2\langle \alpha_\downarrow, \beta_\downarrow \rangle.$$

This proves the first inequality in (28.4). The second follows, in the same way, from the first inequality in (28.2). ■

§29 Partitioned Hermitian matrices

Consider the Hermitian matrices

$$A = \begin{bmatrix} H_1 & E \\ E^* & H_2 \end{bmatrix}, \quad B = \begin{bmatrix} H_1 & O \\ O & H_2 \end{bmatrix}, \quad (29.1)$$

in which the blocks H_1 and H_2 are $m \times m$ and $n \times n$ Hermitian matrices, respectively. From Theorem 8.5 we know that

$$\|\mathrm{Eig}_\downarrow(A) - \mathrm{Eig}_\downarrow(B)\| \;\le\; \|A - B\| = \|E\|. \quad (29.2)$$

It has long been known to numerical analysts that if the spectrum of H_1 is well separated from the spectrum of H_2, then the bound (29.2) can be replaced by one that involves $\|E\|^2$ on the right rather than $\|E\|$.

Such bounds are of interest in the context of computations of eigenvalues. If an approximate eigenspace for a Hermitian operator A has been found, then with respect to this space and its orthogonal complement, A has a block matrix decomposition in which the "residual" E is small. We are interested in knowing the effect of discarding E.

The best result of this kind was published in 2005 by C.-K. Li and R.-C. Li.

Theorem 29.1 Let A and B be Hermitian matrices as in (29.1). Let

$$\eta = d(\operatorname{Spec} H_1, \operatorname{Spec} H_2) = \min\{|\mu_1 - \mu_2| : \mu_1 \in \operatorname{Spec} H_1, \mu_2 \in \operatorname{Spec} H_2\}$$

be the distance between the spectra of H_1 and H_2. Then

$$\|\operatorname{Eig}_\downarrow(A) - \operatorname{Eig}_\downarrow(B)\| \leq \frac{2\|E\|^2}{\eta + \sqrt{\eta^2 + 4\|E\|^2}}. \tag{29.3}$$

Before giving its proof we point out a few salient features of this bound. Let R be the quantity on the right-hand side of (29.3). Then $R \leq \|E\|$, and the two are equal if $\eta = 0$. For large η, (29.3) is a substantial improvement on (29.2). Also we have $R \leq \|E\|^2/\eta$. A bound with this latter quantity instead of R was proved by R. Mathias in 1998.

When $m = n = 1$, the inequality (29.3) can be proved using elementary algebra. Let

$$A = \begin{pmatrix} \alpha & \varepsilon \\ \varepsilon & \beta \end{pmatrix}$$

and assume, without loss of generality, that $\alpha > \beta$. The two eigenvalues of A are

$$\lambda_\pm = \frac{\alpha + \beta \pm \sqrt{(\alpha - \beta)^2 + 4\varepsilon^2}}{2}.$$

The quantities $\lambda_+ - \alpha$ and $\beta - \lambda_-$ are positive and both are equal to

$$\frac{-(\alpha - \beta) + \sqrt{(\alpha - \beta)^2 + 4\varepsilon^2}}{2} = \frac{2\varepsilon^2}{(\alpha - \beta) + \sqrt{(\alpha - \beta)^2 + 4\varepsilon^2}}.$$

So, in this case we even have equality of the two sides of (29.3).

Now let m and n be any two positive integers. Let the eigenvalues of A and B be listed as $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_{m+n}$ and $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_{m+n}$, respectively. We have to prove that

$$|\alpha_j - \beta_j| \leq R \quad \text{for } 1 \leq j \leq m+n. \tag{29.4}$$

Our proof will show a little more. Each β_j is either in $\operatorname{Spec} H_1$ or in $\operatorname{Spec} H_2$. If β_j is in $\operatorname{Spec} H_1$, let $\eta_j = \operatorname{dist}(\beta_j, \operatorname{Spec} H_2)$, and if it is in $\operatorname{Spec} H_2$ let $\eta_j = \operatorname{dist}(\beta_j, \operatorname{Spec} H_1)$. Evidently, $\eta = \min_{1 \leq j \leq m+n} \eta_j$. Let

$$R_j = \frac{2\|E\|^2}{\eta_j + \sqrt{\eta_j^2 + 4\|E\|^2}}.$$

Then $R_j \leq R$. We will show that

$$|\alpha_j - \beta_j| \leq R_j \quad \text{for all } j. \tag{29.5}$$

This implies the inequality (29.4).

We prove (29.5) by induction on $m+n$. We have proved it when $m+n=2$, and assume that it has been proved for Hermitian matrices of size $m+n-1$.

Suppose U and V are unitary matrices such that UH_1U^* and VH_2V^* are diagonal. Then

$$\begin{bmatrix} U & O \\ O & V \end{bmatrix} \begin{bmatrix} H_1 & E \\ E^* & H_2 \end{bmatrix} \begin{bmatrix} U^* & O \\ O & V^* \end{bmatrix} = \begin{bmatrix} UH_1U^* & UEV^* \\ VE^*U^* & VH_2V^* \end{bmatrix}.$$

Since $\|UEV^*\| = \|E\|$, it is no loss of generality if we assume that H_1 and H_2 are diagonal. Further, we can assume that the diagonal entries of A are distinct. The general case follows from this by continuity.

First we prove (29.5) for $j=1$ and $m+n$.

We have assumed that B is diagonal. So β_1 is one of the diagonal entries of B, and we can assume that it is the first one. We know that $\alpha_1 \geq \langle Ae_1, e_1 \rangle = \beta_1$. The matrix

$$A - \alpha_1 I = \begin{bmatrix} H_1 - \alpha_1 I & E \\ E^* & H_2 - \alpha_1 I \end{bmatrix}$$

is congruent to

$$\begin{bmatrix} H_1(\alpha_1) & O \\ O & H_2 - \alpha_1 I \end{bmatrix},$$

where $H_1(\alpha_1) = H_1 - \alpha_1 I - E(H_2 - \alpha_1 I)^{-1}E^*$ is the Schur complement of $H_2 - \alpha_1 I$ in $A - \alpha_1 I$. (See the discussion in §26.) The highest eigenvalue of $A - \alpha_1 I$ is zero. Hence, by Sylvester's law of inertia, the highest eigenvalue of $H_1(\alpha_1)$ is also zero. The highest eigenvalue of $H_1 - \alpha_1 I$ is $\beta_1 - \alpha_1 \leq 0$. So, by Theorem 8.1 applied to the pair $H_1 - \alpha_1 I$ and $H_1(\alpha_1)$ we have

$$\begin{aligned} \alpha_1 - \beta_1 &= |(\beta_1 - \alpha_1) - 0| \leq \|E(H_2 - \alpha_1 I)^{-1}E^*\| \\ &\leq \|E\|^2 \quad \|(H_2 - \alpha_1 I)^{-1}\|. \end{aligned} \tag{29.6}$$

The matrix H_2 is diagonal and all its entries are less than β_1, which in turn is less than α_1. Thus

$$\begin{aligned} \|(H_2 - \alpha_1 I)^{-1}\| &= \left[\min_{\mu \in \operatorname{Spec} H_2} |\alpha_1 - \mu|\right]^{-1} \\ &= \left[(\alpha_1 - \beta_1) + \min_{\mu \in \operatorname{Spec} H_2} (\beta_1 - \mu)\right]^{-1} \\ &= \frac{1}{\delta_1 + \eta_1}, \end{aligned}$$

where $\delta_1 = (\alpha_1 - \beta_1)$. Putting this into (28.10) we get

$$\delta_1 \leq \frac{\|E\|^2}{\delta_1 + \eta_1}.$$

A small calculation shows that this implies

$$\delta_1 \leq \frac{2\|E\|^2}{\eta_1 + \sqrt{\eta_1^2 + 4\|E\|^2}} = R_1 \tag{29.7}$$

which is the inequality we are seeking. We have proved this assuming β_1 is an entry of H_1. The argument can be modified to handle the case when β_1 is an entry of H_2.

We have shown that $|\alpha_1 - \beta_1| \leq R_1$. Applying the same argument to $-A$ in place of A, we see that $|\alpha_{m+n} - \beta_{m+n}| \leq R_{m+n}$.

Now consider any index i, $1 < i < m + n$. If $\alpha_i = \beta_i$, we have nothing to prove. If not, we can assume $\beta_i > \alpha_i$. (Otherwise we replace A by $-A$.) Delete from A the row and column that contain (the diagonal entry) β_n. Let $\widehat{A}$ be the resulting matrix of size $m + n - 1$, and let the eigenvalues of this matrix be enumerated as $\nu_1 \geq \nu_2 \geq \cdots \geq \nu_{m+n-1}$. By Cauchy's interlacing theorem $\alpha_i \geq \nu_i$, and hence, $\beta_i - \alpha_i \leq \beta_i - \nu_i$.

If the diagonal blocks of the truncated matrix $\widehat{A}$ are $\widehat{H}_1$ and $\widehat{H}_2$, then β_i does not belong to one of them, say $\widehat{H}_j$. Let $\widehat{\eta}_i = \text{dist}\left(\beta_i, \widehat{H}_j\right)$. Then $\widehat{\eta}_i \geq \eta_i$. If $\widehat{E}$ is the top right block of $\widehat{A}$. Then $\|\widehat{E}\| \leq \|E\|$.

In order to estimate $|\alpha_i - \beta_i|$, first note that

$$|\alpha_i - \beta_i| = \beta_i - \alpha_i \leq \beta_i - \nu_i.$$

The induction hypothesis says that

$$\beta_i - \nu_i \leq \widehat{R}_i = \frac{2\|\widehat{E}\|^2}{\widehat{\eta}_i + \sqrt{\widehat{\eta}_i^2 + 4\|\widehat{E}\|^2}}.$$

To complete the proof we show that $\widehat{R}_i \leq R_i$. Since $\widehat{\eta}_i \geq \eta_i$ we have

$$\widehat{R}_i \leq \frac{2\|\widehat{E}\|^2}{\eta_i + \sqrt{\eta_i^2 + 4\|\widehat{E}\|^2}} = \frac{\sqrt{\eta_i^2 + 4\|\widehat{E}\|^2} - \eta_i}{2}.$$

Since $\|\widehat{E}\| \leq \|E\|$, this quantity is bounded by

$$\frac{\sqrt{\eta_i^2 + 4\|E\|^2} - \eta_i}{2} = R_i.$$

Thus the inequality (29.5) is valid for all j. ■

§30 Lidskii's Theorem

In Section 9 we derived Lidskii's Theorem from Wielandt's minmax principle. This proof was discovered by Wielandt as he "did not succeed in completing the interesting sketch of a proof given by Lidskii".

Lidskii's Theorem has inspired much research and now many different proofs of it are known. Some of these can be found in the book *Matrix Analysis* by R. Bhatia. A proof much simpler than all others was published by C.-K. Li and R. Mathias in 1999. Their proof is presented in this section.

The theorem says that if A and B are $n \times n$ Hermitian matrices, then for all $k = 1, 2, \ldots, n$ and for all choices of indices $1 \leq i_1 < \cdots < i_k \leq n$ we have

$$\sum_{j=1}^{k} \lambda_{[i_j]}(A+B) \leq \sum_{j=1}^{k} \lambda_{[i_j]}(A) + \sum_{j=1}^{k} \lambda_{[j]}(B). \tag{30.1}$$

Choose and fix k. Replacing B by $B-\lambda_{[k]}(B)I$ does not affect the inequality (30.1) as both sides are diminished by $k\lambda_{[k]}(B)$. So, it is enough to prove the inequality under the assumption that $\lambda_{[k]}(B)=0$.

Let $B = B_+ - B_-$ be the decomposition of B into its positive and negative parts. By this we mean the following. In an appropriate orthonormal basis B can be represented as a diagonal matrix

$$B = \operatorname{diag}(\alpha_1, \ldots, \alpha_p, \alpha_{p+1}, \ldots, \alpha_n)$$

where $\alpha_1, \ldots, \alpha_p$ are nonnegative, and $\alpha_{p+1}, \ldots, \alpha_n$ are negative. In this basis

$$B_+ = \operatorname{diag}(\alpha_1, \ldots, \alpha_p, 0 \ldots, 0)$$

and

$$B_- = \operatorname{diag}(0, \ldots, 0, -\alpha_{p+1}, \ldots, -\alpha_n).$$

Then $B \le B_+$, and by Corollary 6.3 we have

$$\lambda_{[j]}(B) \;\le\; \lambda_{[j]}(B_+) \quad \text{for } 1 \le j \le n.$$

Hence

$$\sum_{j=1}^{k} \left[\lambda_{[i_j]}(A+B) - \lambda_{[i_j]}(A)\right] \;\le\; \sum_{j=1}^{k} \left[\lambda_{[i_j]}(A+B_+) - \lambda_{[i_j]}(A)\right]. \qquad (30.2)$$

By the same argument, since $A \le A + B_+$,

$$\lambda_{[j]}(A) \;\le\; \lambda_{[j]}(A+B_+) \quad \text{for } 1 \le j \le n.$$

So, the sum on the right-hand side of (30.2) is not bigger than

$$\sum_{j=1}^{n} \left[\lambda_{[j]}(A+B_+) - \lambda_{[j]}(A)\right].$$

This last sum is equal to $\operatorname{tr} B_+$, and since $\lambda_{[k]}(B)=0$, this trace is equal to

$$\sum_{j=1}^{k} \lambda_{[j]}(B).$$

We have shown that

$$\sum_{j=1}^{k} \left[\lambda_{[i_j]}(A+B) - \lambda_{[i_j]}(A)\right] \le \sum_{j=1}^{k} \lambda_{[j]}(B).$$

This is the inequality (30.1).

§31 Horn's problem

The Lidskii-Wielandt theorem was the stimulus for a lot of work that culminated in the formulation of a conjecture by Alfred Horn in 1962. Subsequently, it was realised that this conjecture is related to many important questions in diverse areas of mathematics—algebraic geometry, representations of Lie groups, combinatorics, quantum cohomology, and others. One of the most spectacular developments in the last few years is the proof of Horn's conjecture by A. Klyachko, and A. Knutson and T. Tao in papers published in 1998 and 1999, respectively.

The inequalities (8.5) due to H. Weyl, Ky Fan's inequalities in Corollary 6.6, Lidskii's in (30.1) and their generalization due to R. C. Thompson and L. Freede are linear inequalities between eigenvalues of Hermitian matrices A, B, and $A+B$. They have a common feature. Each of them identifies three equinumerous sets of indices I, J, K contained in $\{1, 2, \ldots, n\}$ such that

$$\sum_{k\in K} \lambda_{[k]}(A+B) \ \leq \ \sum_{i\in I} \lambda_{[i]}(A) + \sum_{j\in J} \lambda_{[j]}(B). \qquad (31.1)$$

Horn's problem consists of two questions. Is there a description of all such "admissible triples" (I, J, K)? Are these inequalities sufficient to characterize eigenvalue triples of Hermitian matrices, in that if $\{\alpha_1, \ldots, \alpha_n\}$, $\{\beta_1, \ldots, \beta_n\}$, and $\{\gamma_1, \ldots, \gamma_n\}$ are three n-tuples of real numbers arranged in decreasing order, and

$$\sum_{k\in K} \gamma_k \ \leq \ \sum_{i\in I} \alpha_i + \sum_{j\in J} \beta_j,$$

for all admissible triples (I, J, K), then do there exist Hermitian matrices A and B such that α, β, and γ are the eigenvalues of A, B, and $A + B$, respectively?

A. Horn conjectured that this is so and gave an inductive procedure for describing all admissible triples. We do not discuss this further. There exist several expositions of the problem and its solution from which the reader can learn more.

§32 Lax's theorem and hyperbolic polynomials

Let $\mathcal{R}$ be a real vector space consisting of matrices each of which has purely real eigenvalues. In Section 10 we saw that several eigenvalue inequalities valid for Hermitian matrices A and B are, in fact, valid for pairs of matrices A and B in $\mathcal{R}$.

One such inequality is

$$\lambda_{[1]}(A+B) \leq \lambda_{[1]}(A) + \lambda_{[1]}(B).$$

Applying this to multilinear operators $A^{[k]}$ and $B^{[k]}$ defined in (6.3) we see that

$$\sum_{j=1}^{k} \lambda_{[j]}(A+B) \leq \sum_{j=1}^{k} \lambda_{[j]}(A) + \sum_{j=1}^{k} \lambda_{[j]}(B),$$

for $1 \leq k \leq n$. This is the inequality of Corollary 6.6 for Hermitian matrices. It is natural to wonder whether a Lidskii type inequality (Theorem 9.4) is valid in this context.

Very recently it has been observed that this, and much more, is true. This is a consequence of work on hyperbolic polynomials by J. W. Helton and V. Vinnikov. The application to matrix eigenvalues occurs in papers of L. Gurvits and L. Rodman.

Let $p(\xi, \eta, \lambda)$ be a homogeneous polynomial of degree n in ξ, η, λ such that

(i) the coefficient of λ^n is one, and

(ii) for each fixed real ξ and η, $p(\xi, \eta, \lambda)$ has only real zeros in λ.

Such a polynomial is called *hyperbolic*.

If A and B are $n \times n$ real symmetric matrices, then the polynomial

$$p(\xi, \eta, \lambda) = \det(\xi A + \eta B - \lambda I)$$

is hyperbolic. In his paper in 1958 that we cited in Chapter 3, P. Lax conjectured that, conversely, every hyperbolic polynomial is of this form. This has now been proved.

As a consequence many problems about eigenvalues of matrix pairs A and B in $\mathcal{R}$ have the same answers as for Hermitian pairs. In particular, they satisfy not just Lidskii's inequalities but all of Horn's inequalities as well.

We give pointers to the literature in the Notes at the end. It is worthwhile to record here perturbation bounds in this context.

Let A and B be $n \times n$ matrices such that for all real numbers ξ and η the matrix $\xi A + \eta B$ has purely real eigenvalues. Then we have the majorisations

$$\mathrm{Eig}_{\downarrow}(A) + \mathrm{Eig}_{\uparrow}(B) \prec \mathrm{Eig}(A+B) \prec \mathrm{Eig}_{\downarrow}(A) + \mathrm{Eig}_{\downarrow}(B).$$

From this we can derive, as we did for Hermitian matrices in Chapter 3, the inequalities

$$|||\mathrm{Eig}_{\downarrow}(A) - \mathrm{Eig}_{\downarrow}(B)||| \le |||\mathrm{Eig}(A-B)||| \le |||\mathrm{Eig}_{\downarrow}(A) - \mathrm{Eig}_{\uparrow}(B)||| \tag{32.1}$$

for all unitarily invariant norms.

A word of caution should be injected here. It A and B are Hermitian, then $|||\mathrm{Eig}(A-B)||| = |||A-B|||$. This is not always true in the more general context. From relations between eigenvalues and singular values we have

$$|||\mathrm{Eig}(A-B)||| \le |||A-B|||$$

for all matrices. So, we do get from (32.1) the inequality

$$|||\mathrm{Eig}_{\downarrow}(A) - \mathrm{Eig}_{\downarrow}(B)||| \le |||A-B|||.$$

This is the first inequality of Theorem 9.7. However, the second inequality there

$$|||A-B||| \le |||\mathrm{Eig}_{\downarrow}(A) - \mathrm{Eig}_{\uparrow}(B)|||$$

may not always be true in this context. For example, if $\alpha_1 > \alpha_2$ and

$$A = \begin{bmatrix} \alpha_1 & 1 \\ 0 & \alpha_2 \end{bmatrix}, \quad B = \begin{bmatrix} \alpha_2 & -1 \\ 0 & \alpha_1 \end{bmatrix},$$

then $\mathrm{Eig}_{\downarrow}(A) - \mathrm{Eig}_{\uparrow}(B)$ is the diagonal matrix with entries $\alpha_1 - \alpha_2$ and $\alpha_2 - \alpha_1$ on its diagonal, but $A - B$ has an additional nonzero entry 2 in the top right corner.

Notes and references

K. Löwner states the inequality (28.4) on page 190 of [Lo] and says that it can be established via a simple variational consideration, which can be left to the reader. This seems to have been ignored in much of the literature on perturbation inequalities, even though the paper is extremely famous for other things. We learnt about this from M. Fiedler's commentary on the Hoffman-Wielandt paper, see page 141 of [W]. The great impact of this latter paper can be judged from the fact that the collection [W] contains seven commentaries on it with hardly any overlap between their contents. A. J. Hoffman remarks that "the reason for the theorem's popularity is the publicity given it by Wilkinson in his book". Incidentally, this book does contain a special calculus-based proof of Theorem 28.3.

Theorem 29.1 was proved by C.-K. Li and R.-C. Li [LL]. References to earlier work in this direction can be found in this paper.

The marvelous proof of Lidskii's theorem in Section 30 is taken from [LM2]. Those familiar with the earlier proofs, some of which are given in [B1] will appreciate its simplicity. Yet another proof using nonsmooth analysis is given in [Le]. From Ky Fan's maximum principle (Theorem 6.5) it follow that $\sum_{j=1}^{k} \lambda_{[j]}(A)$ is a convex real valued function on the space of Hermitian matrices, for each $1 \leq k \leq n$. Thus each eigenvalue $\lambda_{[j]}(A)$ is a difference of two convex functions. This simple observation is used to good effect by Hiriart-Urruty and Ye [HY]. The genesis of Horn's problem, its solution, and the connections it has with other areas, have been explained in several expository articles. Two of them are [B3] and [F]. The original papers in which the problem is solved are [Kl] and [KT].

The original work on hyperbolic polynomials that we alluded to in Section 32 is in [V] and [HV]. An excellent article on hyperbolic polynomials and their diverse applications by Bauschke, Güler, Lewis, and Sendov [BGLS] is of special interest for our problems. The authors of [LPR] specifically note that the Lax conjecture follows from the work in [HV]. Several consequences of this for eigenvalues of matrix pairs whose combinations have real roots are derived by Gurvits and Rodman [GR].

9 Spectral variation of normal matrices

§33 The elusive bound

At the time of writing of (the original version of) this book the most intriguing open problem in this subject was whether for any two normal matrices A and B we have the inequality

$$||(\text{Eig}\, A,\ \text{Eig}\, B)|| \leq ||A - B||. \tag{33.1}$$

This was known to be true whenever A and B are Hermitian, when A is Hermitian and B skew Hermitian, when A and B are unitary or are constant multiples of unitaries, when $A, B,$ and $A - B$ are normal, and when A and B are any 2×2 normal matrices.

To the surprise of everyone involved in this problem, John Holbrook discovered a 3×3 counterexample in 1989. Holbrook did this with the assistance of a computer. A bare-hands example was then made up by G. Krause. This is given below. Let

$$\lambda_1 = 1, \quad \lambda_2 = \frac{4 + 5\sqrt{3}i}{13}, \quad \lambda_3 = \frac{-1 + 2\sqrt{3}i}{13};$$

let v be the vector

$$v = \begin{pmatrix} \sqrt{5/8} \\ 1/2 \\ \sqrt{1/8} \end{pmatrix},$$

and U the 3×3 unitary matrix $I - 2vv^*$. Let $A = \text{diag}(\lambda_1, \lambda_2, \lambda_3)$ and $B = -U^*AU$. Then A and B are normal, and the eigenvalues of B are the negatives of the eigenvalues of A. A calculation shows

$$||(\text{Eig}\, A,\ \text{Eig}\, B)|| = \sqrt{\frac{28}{13}}, \quad ||A - B|| = \sqrt{\frac{27}{13}}.$$

So, in this example

$$||(\operatorname{Eig} A,\ \operatorname{Eig} B)|| > 1.0183\ ||A - B||.$$

In Section 16 we proved that there exists a number c such that

$$||(\operatorname{Eig} A,\ \operatorname{Eig} B)|| \leq c\ ||A - B||, \tag{33.2}$$

and that $c \leq c_2$, where c_2 is defined in (16.8). The extremal problem leading to c_2 was reformulated by R. Bhatia, C. Davis and P. Koosis and it was shown that

$$c_2 = \inf \left\{ \int_0^\infty |\widehat{g}(t)|\, dt : g \,\text{even}, \operatorname{supp} g = [-1, 1], \int_{-1}^1 g = 1, \widehat{g} \in L_1 \right\}. \tag{33.3}$$

An obvious choice of a function g in the class above is $g(t) = 1 - |t|$. In this case

$$\widehat{g}(t) = \frac{\sin^2(t/2)}{(t/2)^2},$$

and, therefore $c_2 \leq \pi$. Choosing another function g Bhatia, Davis, and Koosis showed that

$$c_2 \ \leq\ \frac{\pi}{2} \int_0^\pi \frac{\sin\, t}{t} dt \ <\ 2.90901. \tag{33.4}$$

It turns out that this bound is very nearly the best for c_2. The extremal problem (16.8) occurs in an entirely different context in the work of L. Hörmander and B. Bernhardsson. They also show the equivalence of that problem and (33.3) and then go on to show that

$$2.903887282 < c_2 < 2.90388728275228.$$

The only method for finding *any* bound for c in (33.2) that has been successful so far is the one that shows $c \leq c_2$. The discussion above shows that no further significant improvement on estimating c is possible via this argument.

In Section 13 we introduced a different method for estimating $||(\operatorname{Eig} A, \operatorname{Eig} B)||$ that had some success: it led to the inequality (33.1) in special cases covered by Theorems 13.3 and 13.5. These considerations lead to an interesting question.

Let $\mathcal{N}(n)$ be the set consisting of all $n \times n$ normal matrices. For any two elements of $\mathcal{N}(n)$ let γ be a rectifiable normal path joining A and B and let

$$\ell(A, B) \ =\ \inf\ \ell_{||\cdot||}(\gamma)$$

be the infimum of the lengths of all such paths measured in the norm $||\cdot||$. We conjecture that there exists a number $k(n)$ such that

$$\ell(A,B) \ \leq \ k(n)\ ||A-B|| \quad \text{for all } A,B \in \mathcal{N}(n). \tag{33.5}$$

Theorem 13.7 shows that $k(2) = 1$, and $k(3) > 1$. Let $k = \sup_n k(n)$. With no evidence to support us, we do not conjecture a value for k, but we will not be surprised if it turns out that $k = \pi/2$.

We do know from our discussion in Section 13 that the inequality (33.2) is valid with $c = k$. There could well be better ways of finding the best c.

The problem of finding an upper bound for $||A - B||$ akin to the one in (8.8) is not difficult. If A and B are normal matrices with eigenvalues $\{\alpha_1, \ldots, \alpha_n\}$ and $\{\beta_1, \ldots, \beta_n\}$, respectively, then there exists a permutation σ such that

$$||A-B|| \ \leq \ \sqrt{2} \max_{1\leq j\leq n} |\alpha_j - \beta_{\sigma(j)}|. \tag{33.6}$$

One proof of this can be based on the fact that there exists a point γ in the plane such that

$$\max_i |\alpha_i - \gamma| \ + \ \max_j |\beta_j - \gamma| \ \leq \ \sqrt{2} \max_{i,j} |\alpha_i - \beta_j|.$$

The 2×2 example on Page 86 shows that the inequality (33.6) is best possible.

§34 Hermitian and skew-Hermitian matrices

In Section 18 we proved an inequality for the distance between the eigenvalues of a Hermitian and a skew-Hermitian matrix, and on Page 119 mentioned some generalizations (then new) and a conjecture (since then proved by X. Zhan). This case is not of much practical interest since a small perturbation will either change a Hermitian matrix to an arbitrary one, or (if it is a structured perturbation) keep it Hermitian. However, the study of this special case reveals several interesting phenomena of the general perturbation problem, involves very interesting matrix techniques, leads to striking results, and is connected to other important problems. Therefore, we present it here in brief.

Let $\{x_j\}_j$ stand for an n-vector whose jth component is x_j.

Theorem 34.1 Let A and B be Hermitian matrices with eigenvalues α_j and β_j, respectively, ordered so that

$$|\alpha_1| \geq \cdots \geq |\alpha_n| \quad \text{and} \quad |\beta_1| \geq \cdots \geq |\beta_n|.$$

Let s_j be the singular values of $T = A + iB$. Then we have the majorization relations

$$\left\{|\alpha_j + i\beta_{n-j+1}|^2\right\}_j \prec \left\{s_j^2\right\}_j, \tag{34.1}$$

$$\left\{\frac{s_j^2 + s_{n-j+1}^2}{2}\right\}_j \prec \left\{|\alpha_j + i\beta_j|^2\right\}_j. \tag{34.2}$$

Proof For any two Hermitian matrices X and Y we have the majorizations (proved in Section 9):

$$\mathrm{Eig}_\downarrow(X) + \mathrm{Eig}_\uparrow(Y) \prec \mathrm{Eig}(X+Y) \prec \mathrm{Eig}_\downarrow(X) + \mathrm{Eig}_\downarrow(Y). \tag{34.3}$$

If $X = A^2$ and $Y = B^2$, this gives

$$\left\{|\alpha_j + i\beta_{n-j+1}|^2\right\}_j \prec \left\{s_j(A^2+B^2)\right\}_j \prec \left\{|\alpha_j + i\beta_j|^2\right\}_j. \tag{34.4}$$

If we choose $X = T^*T/2$, $Y = TT^*/2$, use the identity

$$\frac{T^*T + TT^*}{2} = A^2 + B^2,$$

and the fact $s_j(T^*T) = s_j(TT^*) = s_j^2$, then from the first majorization in (34.3) we get

$$\left\{\frac{s_j^2 + s_{n-j+1}^2}{2}\right\}_j \prec \left\{s_j(A^2+B^2)\right\}_j \prec \left\{s_j^2\right\}_j. \tag{34.5}$$

The two assertions of the theorem follow from (34.4) and (34.5). ■

The two relations (34.1) and (34.2) contain a wealth of information. The function $g(t) = t^{p/2}$ is convex on $[0, \infty)$ when $p \geq 2$. So, using Corollary 3.4, we obtain from (34.1) the weak majorization

$$\left\{|\alpha_j + i\beta_{n-j+1}|^p\right\}_j \prec_w \left\{s_j^p\right\}_j \quad \text{for } 2 \leq p. \tag{34.6}$$

This implies several inequalities, one of which says

$$\sum_{j=1}^{n} |\alpha_j + i\beta_{n-j+1}|^p \;\le\; \sum_{j=1}^{n} s_j^p \quad \text{for } 2 \le p. \tag{34.7}$$

For $1 \le p \le 2$, the function $g(t) = t^{p/2}$ is concave, and the inequality (34.7) is reversed.

The same arguments applied to (34.2) lead to the inequality

$$\frac{1}{2^{p/2}} \sum_{j=1}^{n} (s_j^2 + s_{n-j+1}^2)^{p/2} \;\le\; \sum_{j=1}^{n} |\alpha_j + i\beta_j|^p \quad \text{for } 2 \le p. \tag{34.8}$$

The inequality is reversed for $1 \le p \le 2$.

For fixed nonnegative real numbers a_1 and a_2 the function $(a_1^t + a_2^t)^{1/t}$ is a monotonically decreasing function of t on $(0,\infty)$. Hence for $p \ge 2$

$$s_j^p + s_{n-j+1}^p \;\le\; \left(s_j^2 + s_{n-j+1}^2\right)^{p/2},$$

and the inequality is reversed for $1 \le p \le 2$. Hence, from (34.8) we obtain

$$2^{1-p/2} \sum_{j=1}^{n} s_j^p \;\le\; \sum_{j=1}^{n} |\alpha_j + i\beta_j|^p \quad \text{for } 2 \le p. \tag{34.9}$$

The inequality is reversed for $1 \le p \le 2$.

These inequalities can be interpreted as perturbation bounds analogous to results for Hermitian matrices in Chapter 3. Let us denote by $\mathrm{Eig}_{|\downarrow|}(A)$ the n-tuple of eigenvalues of A arranged in decreasing order of their absolute values, and by $\mathrm{Eig}_{|\uparrow|}(A)$ the same n-tuple arranged in increasing order of absolute values. With an obvious change of notation, the inequalities (34.8) and (34.9) can be stated as follows.

Theorem 34.2 Let A be a Hermitian and B a skew-Hermitian matrix. Then for $2 \le p \le \infty$, we have

$$\begin{aligned} \|\mathrm{Eig}_{|\downarrow|}(A) - \mathrm{Eig}_{|\uparrow|}(B)\| &\le \|A - B\|_p, \\ \|A - B\|_p &\le 2^{1/2-1/p}\|\mathrm{Eig}_{|\downarrow|}(A) - \mathrm{Eig}_{|\downarrow|}(B)\|_p, \end{aligned}$$

and for $1 \le p \le 2$ we have

$$\begin{aligned} \|\mathrm{Eig}_{|\downarrow|}(A) - \mathrm{Eig}_{|\downarrow|}(B)\| &\le 2^{1/p-1/2}\|A - B\|_p, \\ \|A - B\|_p &\le \|\mathrm{Eig}_{|\downarrow|}(A) - \mathrm{Eig}_{|\uparrow|}(B)\|. \end{aligned}$$

The 2×2 example on Page 86 shows that all these inequalities are sharp.

An interesting feature of these inequalities is that the optimal matching of eigenvalues—the permutation W that minimises

$$\|D(A) - WD(B)W^{-1}\|_p$$

in (11.3), is different for $2 \le p$ and $1 \le p \le 2$. In fact the best matching in one case is the worst in the other. This is quite different from the behaviour of Hermitian pairs A and B, where the same matching is optimal for all unitarily-invariant norms.

For the sake of completeness, it is desirable to have bounds for the whole family of unitarily-invariant norms in this case as well. The upper bound for $|||A-B|||$ comes cheap. We have

$$\begin{aligned}\sum_{j=1}^{k} s_j(A-B) &\le \sum_{j=1}^{k} [s_j(A)+s_j(B)] = \sum_{j=1}^{k} [|\alpha_j|+|\beta_j|] \\ &\le \sum_{j=1}^{k} \sqrt{2}\,(\alpha_j^2+\beta_j^2)^{1/2} = \sum_{j=1}^{k} \sqrt{2}\,|\alpha_j - i\beta_j|.\end{aligned}$$

Thus, when A is Hermitian and B skew-Hermitian, we have

$$|||A-B||| \le \sqrt{2}\,|||\mathrm{Eig}_{|\downarrow|}(A) - \mathrm{Eig}_{|\downarrow|}(B)||| \tag{34.10}$$

for all unitarily-invariant norms. This inequality is sharp for the operator norm $\|\cdot\|$.

To prove the complementary lower bound we need a little more intricate argument. Switch to the notations of Theorem 34.1 for convenience. We will prove that

$$\{|\alpha_j + i\beta_j|\}_j \prec_w \sqrt{2}\,\{s_j\}_j, \tag{34.11}$$

or, in other words,

$$|||\mathrm{diag}(\alpha_1+\beta_1,\ldots,\alpha_n+\beta_n)||| \le \sqrt{2}\,|||T|||. \tag{34.12}$$

We will use the characterisation of Ky Fan norms given in Theorem 27.2.

Since $|\alpha_1|$ and $|\beta_1|$ are bounded by s_1, we have $|\alpha_1+i\beta_1| \le \sqrt{2}\,s_1$. So the inequality (34.12) is true for the norm $\|\cdot\|$. It is true also for the norm $\|\cdot\|_{\mathrm{tr}}$ by the remark following (34.9).

Now fix k, $1 < k < n$. By Theorem 27.2 there exist matrices X and Y such that $T = X + Y$ and $\|T\|_k = \|X\|_{\text{tr}} + k\|Y\|$. Let C, D, E, F be Hermitian matrices such that $X = C + iD$ and $Y = E + iF$. Then $A = C + E$, and $B = D + F$. Thus

$$\|\text{diag}(\alpha_1 + \beta_1, \ldots, \alpha_k + \beta_k)\|_k = \sum_{j=1}^{k} |s_j(C+E) + i\, s_j(D+F)|.$$

By Corollary 6.2, $s_j(C+E) \le s_j(C) + s_1(E)$ and $s_j(D+F) \le s_j(D) + s_1(F)$. So the last sum above is bounded by

$$\begin{aligned}
&\sum_{j=1}^{k} |s_j(C) + s_1(E) + i(s_j(D) + s_1(F))| \\
&\quad \le \sum_{j=1}^{k} |s_j(C) + i\, s_j(D)| + k\ |s_1(E) + i\, s_1(F)| \\
&\quad \le \sum_{j=1}^{n} |s_j(C) + i\, s_j(D)| + k\,|s_1(E) + i\, s_1(F)|\,.
\end{aligned}$$

We have observed that the inequality (34.12) is true for the norms $\|\cdot\|_{\text{tr}}$ and $\|\cdot\|$. So, this expression is bounded by

$$\sqrt{2}\ \|X\|_{\text{tr}} + \sqrt{2}\ k\|Y\| = \sqrt{2}\ |||T|||_k.$$

This proves (34.12).

With a change of notation, this says that if A is Hermitian and B skew-Hermitian, then for all unitarily-invariant norms

$$|||\text{Eig}_{|\downarrow|}(A) - \text{Eig}_{|\downarrow|}(B)||| \ \le\ \sqrt{2}\ |||A - B|||. \tag{34.13}$$

This inequality complements (34.10), and is sharp for the norm $\|\cdot\|_1 = \|\cdot\|_{\text{tr}}$.

Finally we remark that some of the majorizations in this section involve squares of vectors. They lead to inequalities for Q-norms that we introduced in Section 27.

§35 A table of known and unknown constants

We are interested in finding the best constants in inequalities of the form

$$\|\,(\text{Eig}\,A,\ \text{Eig}\,B)\,\|_p \ \le\ c\,(p)\|A - B\|_p, \tag{35.1}$$

where A and B are $n \times n$ matrices in some subclass of normal matrices, and the constant $c(p)$ depends on the Schatten norm $\|\cdot\|_p$ but is independent of n. It is clear that we must have $c(p) \geq 1$ in all cases. The following table shows what is known on this problem. As usual, $\|\cdot\|_\infty$ stands for the operator norm.

	$p=1$	$1<p<2$	$p=2$	$2<p<\infty$	$p=\infty$	all $\|\|\|\cdot\|\|\|$
A, B Hermitian	1	1	1	1	1	1
$A=A^*$, $B=-B^*$	$\sqrt{2}$	$2^{1/p-1/2}$	1	1	1	$\sqrt{2}$
A, B unitary	$\pi/2$	?	1	?	1	$\pi/2$
A, B normal	?	?	1	?	2.91	?

The table should be interpreted as follows. The entry 1, wherever it occurs, is the sharp value of $c(p)$ in (35.1) for the special case under consideration. In the second row the constant $\sqrt{2}$ is best possible for $p=1$, and it works for all unitarily-invariant norms. For other p the best possible values are as indicated. In the third row $\pi/2$ is the best possible for $p=1$ and works for all unitarily-invariant norms. The question marks indicate that the best constants are not known for these cases. The number 2.91 in the last row is an upper bound for the best constant. From the data in the first two rows of the table it is reasonable to conjecture that

(i) For A, B unitary, $c(p)=1$ for all $2 \leq p \leq \infty$.

(ii) For A, B unrestricted normal matrices, $c(p)=1$ *only* in the case $p=2$. (We do know that $c(p)>1$ for $1 \leq p < 2$ and for $p=\infty$. From the latter it follows that there exist large values of p for which $c(p)>1$.)

Notes and references

Holbrook's example showing that the inequality (33.1) is not true for all normal matrices appears in [Ho]. The bound (33.4) is established in [BDKo], and the narrowing down mentioned after that in [HB]. The proof of (33.6) and its generalization (42.6) may be found in [BES].

Theorems 34.1 and 34.2 were proved by T. Ando and R. Bhatia [AB]. The bound (34.11) was conjectured there and later proved by X. Zhan

[Z]. Let $T = A + iB$ be the decomposition of any operator T into its Hermitian and skew-Hermitian parts. Then $\|T\|_F^2 = \|A\|_F^2 + \|B\|_F^2$. For norms other than the Frobenius norm, interesting inequalities relating the three operators $T, A,$ and B can be found in [BK], [BZ1] and [BZ2].

10 Spectral variation of diagonalizable matrices

Some of the inequalities in Chapters 3 and 4 can be generalized to matrices that are *diagonalizable* (similar to diagonal matrices). Normal matrices are special in being diagonalizable via unitary similarities. This feature is highlighted in these more general versions.

§36 The Sylvester equation and commutator inequalities

In Section 16 we saw that if A and B are any two matrices whose spectra are disjoint from each other, then for any S the Sylvester equation

$$AQ - QB = S \tag{36.1}$$

has a unique solution Q. For normal A and B we obtained a bound for $\|Q\|$ in Theorem 16.3. This bound can be improved if the spectra of A and B are separated in special ways.

Theorem 36.1 Let A and B be any two matrices whose spectra are contained in the open right-half plane and the open left-half plane, respectively. Then the solution of the equation (36.1) can be expressed as

$$Q = \int_0^\infty e^{-tA}\, S\, e^{tB} dt. \tag{36.2}$$

Proof The hypotheses on A and B ensure that the integral is convergent. If Q is given by this formula, then

$$\begin{aligned} AQ - QB &= \int_0^\infty \left(A\,e^{-tA}\,S\,e^{tB} - e^{-tA}\,S\,e^{tB}B\right) dt \\ &= e^{-tA}\,S\,e^{tB}\Big|_0^\infty = S. \end{aligned}$$

Thus Q satisfies the equation (36.1). ■

Corollary 36.2 Let A and B be normal matrices whose spectra are contained in half-planes separated by distance δ. Then the solution of (36.1) satisfies the inequality

$$|||Q||| \leq \frac{1}{\delta}|||S|||. \tag{36.3}$$

Proof By applying a rotation and translation, we may assume that $\operatorname{Spec} A$ and $\operatorname{Spec} B$ are subsets of the half-planes $\operatorname{Re} z > \delta/2$ and $\operatorname{Re} z < -\delta/2$, respectively. Then for $t > 0$, $\|e^{-tA}\|$ and $\|e^{tB}\|$ are bounded by $e^{-t\delta/2}$. So, from (36.2) we have, using Proposition 7.7

$$\begin{aligned} |||Q||| &\leq \int_0^\infty \|e^{-tA}\|\ |||S|||\ \|e^{tB}\| dt \\ &\leq \int_0^\infty e^{-t\delta} dt\ |||S||| \\ &= \frac{1}{\delta}\ |||S|||. \end{aligned}$$ ■

We use this to estimate norms of operators of the form $A\Gamma - \Gamma B$. These are called generalised commutators.

Theorem 36.3 Let A and B be Hermitian matrices and let $\Gamma \geq \gamma I > 0$ (a positive definite matrix with smallest eigenvalue γ). Then

$$|||A\Gamma - \Gamma B||| \geq \gamma\ |||A - B|||. \tag{36.4}$$

Proof Let $T = A\Gamma - \Gamma B$ and $S = T^* + T$. Then

$$S = \Gamma(A - B) + (A - B)\Gamma.$$

This is an equation of the type (36.1), and Corollary 36.2 shows that

$$|||A - B||| \leq \frac{1}{2\gamma}|||S||| \leq \frac{1}{\gamma}|||T||| = \frac{1}{\gamma}|||A\Gamma - \Gamma B|||.$$ ■

Now if Γ is as above but A and B are any two matrices, then applying the Theorem to the Hermitian matrices $\begin{bmatrix} O & A \\ A^* & O \end{bmatrix}$, $\begin{bmatrix} O & B \\ B^* & O \end{bmatrix}$, and to $\begin{bmatrix} \Gamma & O \\ O & \Gamma \end{bmatrix}$ we obtain

$$\left|\left|\left| \begin{bmatrix} A\Gamma - \Gamma B & O \\ O & A^*\Gamma - \Gamma B^* \end{bmatrix} \right|\right|\right| \geq \gamma \left|\left|\left| \begin{bmatrix} A - B & O \\ O & A^* - B^* \end{bmatrix} \right|\right|\right|. \quad (36.5)$$

The matrix $A^* - B^*$ has the same singular values as $A - B$. If A and B are unitary, then

$$\begin{aligned} s_j(A^*\Gamma - \Gamma B^*) &= s_j(A(A^*\Gamma - \Gamma B^*)B) = s_j(\Gamma B - A\Gamma) \\ &= s_j(A\Gamma - \Gamma B). \end{aligned}$$

So in this case we see from (36.5) that $\|A\Gamma - \Gamma B\|_k \geq \gamma \|A - B\|_k$ for all Ky Fan norms, and hence we have:

Corollary 36.4 If A and B are unitary, and $\Gamma \geq \gamma I > 0$, then

$$|||A\Gamma - \Gamma B||| \geq \gamma\, |||A - B|||.$$

Next consider the case when A and B are normal. If $\alpha_1, \ldots, \alpha_n$ are the eigenvalues of A, then using an orthonormal basis of eigenvectors, one sees that for every X we have

$$\|AX - XA\|_F^2 = \sum_{i,j} |\alpha_i - \alpha_j|^2 \, |x_{ij}|^2 = \|A^*X - XA^*\|_F^2.$$

Applying this to $\begin{bmatrix} A & O \\ O & B \end{bmatrix}$ in place of A, and $\begin{bmatrix} O & X \\ O & O \end{bmatrix}$ in place of X, we see that

$$\|AX - XB\|_F^2 = \|A^*X - XB^*\|_F^2. \quad (36.6)$$

Hence, we have from (36.5):

Corollary 36.5 If A and B are normal, and $\Gamma \geq \gamma I > O$, then

$$\|A\Gamma - \Gamma B\|_F \geq \gamma\, \|A - B\|_F. \quad (36.7)$$

In the case of normal matrices we have to be content with the special Frobenius norm, as there are examples that show (36.7) is not always valid for the operator norm $\|\cdot\|$.

§37 Diagonalizable matrices

In this section we consider matrices

$$A = SD_1S^{-1} \quad \text{and} \quad B = TD_2T^{-1}, \tag{37.1}$$

where D_1 and D_2 are diagonal, and S and T are invertible. The *condition number* of S is defined as

$$c(S) = \|S\| \, \|S^{-1}\|.$$

We prove three theorems covering the cases when the spectra of A and B are real, lie on the unit circle, or are unrestricted. When the matrices S and T are unitary, their condition numbers are equal to one, and our results reduce to those proved in Chapters 3 and 4 for Hermitian, unitary, and normal matrices.

Theorem 37.1 Suppose A and B are as in (36.8) and assume further that D_1 and D_2 are real. Then

$$|||\mathrm{Eig}_\downarrow(A) - \mathrm{Eig}_\downarrow(B)||| \;\le\; \sqrt{c(S)c(T)} \; |||A - B||| \tag{37.2}$$

for every unitarily-invariant norm.

Proof From the equations

$$A - B = S(D_1S^{-1}T - S^{-1}TD_2)T^{-1},$$

and

$$A - B = T(T^{-1}SD_1 - D_2T^{-1}S)S^{-1},$$

we get the inequalities

$$|||D_1S^{-1}T - S^{-1}TD_2||| \;\le\; |||S^{-1}(A-B)T||| \;\le\; ||S^{-1}|| \; |||A - B||| \; ||T||,$$

and

$$|||T^{-1}SD_1 - D_2T^{-1}S||| \;\le\; |||T^{-1}(A-B)S||| \;\le\; ||T^{-1}|| \; |||A - B||| \; ||S||.$$

Let $S^{-1}T = U\Gamma V$ be the singular value decomposition. Then

$$\begin{aligned} |||D_1S^{-1}T - S^{-1}TD_2||| &= |||D_1U\Gamma V - U\Gamma VD_2||| \\ &= |||U^*D_1U\Gamma - \Gamma VD_2V^*||| \\ &= |||A'\Gamma - \Gamma B'|||, \end{aligned}$$

where $A' = U^*D_1U$ and $B' = VD_2V^*$ are Hermitian matrices. The same argument, with the relation $T^{-1}S = V^*\Gamma^{-1}U^*$, shows

$$|||T^{-1}SD_1 - D_2T^{-1}S||| = |||\Gamma^{-1}A' - B'\Gamma^{-1}|||.$$

So, the two inequalities above can be expressed as

$$\alpha|||A - B||| \geq |||A'\Gamma - \Gamma B'|||,$$

and

$$\beta|||A - B||| \geq |||A'\Gamma^{-1} - \Gamma^{-1}B'|||,$$

where $\alpha = ||S^{-1}||\ ||T||$, and $\beta = ||T^{-1}||\ ||S||$. The last two inequalities, combined with the triangle inequality, give

$$2\,|||A - B||| \geq |||A'\left(\frac{\Gamma}{\alpha} + \frac{\Gamma^{-1}}{\beta}\right) - \left(\frac{\Gamma}{\alpha} + \frac{\Gamma^{-1}}{\beta}\right)B'|||.$$

The arithmetic-geometric mean inequality implies that

$$\frac{\Gamma}{\alpha} + \frac{\Gamma^{-1}}{\beta} \geq \frac{2}{\sqrt{\alpha\beta}}\,I.$$

Theorem 36.3 can now be applied to get

$$2\,|||A - B||| \geq \frac{2}{\sqrt{\alpha\beta}}\,|||A' - B'|||.$$

The matrices A' and B' are Hermitian and have the same eigenvalues as those of A and B. So using Theorem 9.7 we get

$$|||A - B||| \geq \frac{1}{\sqrt{\alpha\beta}}\,|||\mathrm{Eig}_{\downarrow}(A) - \mathrm{Eig}_{\downarrow}(B)|||.$$

This is the desired inequality (37.2). ■

Exactly the same arguments combining Corollary 36.4 with Theorems 13.6 and 14.7, and Corollary 36.5 with Theorem 15.1 give the following.

Theorem 37.2 Let A and B be as in (37.1) and assume in addition that all eigenvalues of A and B have modulus 1. Then

$$||(\mathrm{Eig}\,A,\ \mathrm{Eig}\,B)|| \leq \sqrt{c(S)c(T)}\ ||A - B||,$$

and

$$|||(\mathrm{Eig}\,A,\ \mathrm{Eig}\,B)||| \leq \frac{\pi}{2}\sqrt{c(S)c(T)}\ |||A - B|||,$$

for all unitarily-invariant norms.

Theorem 37.3 Let A and B be any two matrices as in (37.1). Then

$$||(\mathrm{Eig}\, A,\ \mathrm{Eig}\, B)||_F \ \leq\ \sqrt{c(S)c(T)}\ ||A - B||_F.$$

J. H. Wilkinson (*The Algebraic Eigenvalue Problem*, p.87) remarks that the overall sensitivity of the eigenvalues of a diagonalizable matrix A is dependent on the size of $c(S)$ which may be regarded as a condition number of A with respect to its eigenvalue problem. Theorem 25.1 due to Bauer and Fike is one manifestation of this. Results of this Section carry this line of thinking further and bring out the special role that normality plays in controlling the behaviour of eigenvalues under perturbations. Chapter 4 of Wilkinson's book contains an illuminating discussion of condition numbers.

Notes and references

The Sylvester equation (36.1) arises in diverse contexts. An expository article on this theme is [BR]. The ideas behind the theorems in Sections 36 and 37 go back to W. Kahan [K]. The inequality (36.4) for the special operator norm was proved by Kahan, and was the motivation for J.-G. Sun [S1] to prove (36.7). This, in turn, provided the motivation for the generalization (36.4) in [BDKi]. All these authors applied their results to derive the perturbation bounds of Section 37 in a weaker form. Their versions of the inequalities had the factors $c(S)c(T)$ without the square root. The inequality (37.2) for the operator norm alone was proved by T.-X. Lu [Lu]. All the theorems of Section 37 are due to R. Bhatia, F. Kittaneh, and R.-C. Li [BKL].

11 The general spectral variation problem

The bounds (19.6) and (20.8) for the optimal matching distance between the roots of two polynomials of degree n, and for the eigenvalues of two $n \times n$ matrices, contain a factor $c(n)$. The best value of $c(n)$ known at the time of writing the original version was n. Since then significantly stronger bounds have been established with $c(n) < 4$. We present these and draw attention to the problems that remain open.

§38 The distance between roots of polynomials

A well-known theorem of Chebyshev, used frequently in approximation theory, says that if p is a monic polynomial of degree n, then

$$\max_{0 \le t \le 1} |p(t)| \ge \frac{1}{2^{2n-1}}. \tag{38.1}$$

We need a small extension of this:

Lemma 38.1 Let C be a continuous curve in the complex plane with endpoints a and b. Then for every monic polynomial of degree n

$$\max_{\lambda \in C} |p(\lambda)| \ge \frac{|b-a|^n}{2^{2n-1}}. \tag{38.2}$$

Proof Let L be the straight line through the points a and b, and S the segment between a and b:

$$\begin{aligned} L &= \{z : z = a + t(b-a), \quad t \in \mathbb{R}\}, \\ S &= \{z : z = a + t(b-a), \quad 0 \le t \le 1\}. \end{aligned}$$

For $z \in \mathbb{C}$ let z' be its orthogonal projection onto L. Then $|z - w| \geq |z' - w'|$ for all $z, w \in \mathbb{C}$.

Let $\lambda_1, \ldots, \lambda_n$ be the roots of p, and let $\lambda_i' = a + t_i(b - a)$ be their projections onto L. If $z = a + t(b - a)$ is any point on L, then

$$\prod_{i=1}^{n} |z - \lambda_i'| = \prod_{i=1}^{n} |(t - t_i)(b - a)| = |b - a|^n \prod_{i=1}^{n} |t - t_i|.$$

The inequality (38.1), applied to the polynomial $\prod_{i=1}^{n} (t - t_i)$, shows that there exists a point z_0 on S such that

$$\prod_{i=1}^{n} |z_0 - \lambda_i'| \geq \frac{|b - a|^n}{2^{2n-1}}.$$

The point z_0 is the projection onto L of some point λ_0 on C; i.e. $z_0 = \lambda_0'$. Since $|\lambda_0 - \lambda_i| \geq |\lambda_0' - \lambda_i'|$, the inequality displayed above shows that

$$\prod_{i=1}^{n} |\lambda_0 - \lambda_i| \geq \frac{|b - a|^n}{2^{2n-1}}.$$

This proves (38.2). ■

Let f and g be monic polynomials as in (19.1) with roots $\alpha_1, \ldots, \alpha_n$, and $\beta_1, \ldots, \beta_n$, respectively. As mentioned in Remark 19.2, we have $\max_j |\alpha_j| \leq 2 \max_j |a_j|^{1/j}$. Let $\Gamma = \max_j \left(|a_j|^{1/j}, |b_j|^{1/j}\right)$ and let

$$\gamma = 2\Gamma. \tag{38.3}$$

(This is different from the definition in (19.2), and the change is needed for our argument.). Let Θ be the quantity defined by (19.3) with this new definition of γ. We have the following significant improvement of Theorem 19.4.

Theorem 38.2 The optimal matching distance between the n-tuples $\operatorname{Root} f$ and $\operatorname{Root} g$ is bounded as

$$d(\operatorname{Root} f, \operatorname{Root} g) < 4\,\Theta. \tag{38.4}$$

Proof Let $g_t = (1 - t)f + tg$, $0 \leq t \leq 1$. Then g_t is a family of monic polynomials. If λ is a root of any g_t, then $|\lambda| \leq \gamma$. Since $g_t(\lambda) = 0$, we have

$$\begin{aligned} |f(\lambda)| &= |t[f(\lambda)-g(\lambda)]| \leq |f(\lambda)-g(\lambda)| \\ &\leq \sum_{k=1}^{n} |a_k - b_k| \, |\lambda|^{n-k} \leq \Theta^n. \end{aligned}$$

As t changes from 0 to 1, the roots of g_t trace out n continuous curves in the plane starting at the roots of f and ending at the roots of g. Let C be any one of these curves, and let a and b be its endpoints. Then by Lemma 38.1 there exists a point λ on C such that

$$|f(\lambda)| \geq \frac{|b-a|^n}{2^{2n-1}}.$$

The two inequalities displayed above show that

$$|a-b| \leq 4 \cdot 2^{-1/n}\, \Theta.$$

This implies that (38.4) is true. ■

How sharp is this inequality? For each positive integer n, let

$$c(n) = \sup\left\{ \frac{d(\text{Root } f, \text{Root } g)}{\Theta(f,g)} : f, g \text{ monic polynomials of degree } n \right\}, \tag{38.5}$$

where $\Theta(f,g)$ is the quantity defined by (19.3). We have shown that

$$\sup_n c(n) \leq 4. \tag{38.6}$$

It is known that

$$\sup_n c(n) \geq 2. \tag{38.7}$$

This follows from an example constructed in a paper of R. Bhatia, L. Elsner, and G. Krause. We conjecture that

$$\sup_n c(n) = 2.$$

G. Krause has shown that $\sup_n c(n) \leq 3.08$.

§39 The distance between eigenvalues

In Section 20 we derived the general spectral variation bound (20.8) using the inequality (19.6) for roots of polynomials. Since the factor $c(n)$ in the latter can be replaced by 4, the same improvement of (20.8) follows as a consequence. However, it is possible to prove a slightly better inequality by a simpler argument that avoids the use of characteristic polynomial.

Let X and Y be any two $n\times n$ matrices, and let λ be an eigenvalue of Y. The argument with Hadamard's inequality on Page 120 shows that

$$|\det(X-\lambda I)| \leq \|X-Y\| \, (\|X\|+\|Y\|)^{n-1}. \tag{39.1}$$

Theorem 39.1 Let A and B be any two $n\times n$ matrices. Then

$$d(\operatorname{Eig} A,\ \operatorname{Eig} B) \leq 4(\|A\|+\|B\|)^{1-1/n}\,\|A-B\|^{1/n} \tag{39.2}$$

Proof The proof is very similar to that of Theorem 38.2. Let $A(t) = (1-t)A+tB$, $0\leq t\leq 1$. As t changes from 0 to 1, the eigenvalues of $A(t)$ trace out n curves in the plane starting at the eigenvalues of A and ending at the eigenvalues of B. So, to prove (39.2) it suffices to show that if C is one of these curves and a, b are its endpoints, then $|a-b|$ is bounded by the right-hand side of (39.2).

Assume, without loss of generality, that $\|A\|\leq\|B\|$. Then $\|A(t)\|\leq \|B\|$ for $0\leq t\leq 1$. By Lemma 38.1, there exists a point λ_0 on C such that

$$|\det(A-\lambda_0 I)| \geq \frac{|b-a|^n}{2^{2n-1}}.$$

The point λ_0 is an eigenvalue of some matrix $A(t_0)$. The inequality (39.1) shows that

$$\begin{aligned}|\det(A-\lambda_0 I| &\leq \|A-A(t)\|\,(\|A\|+\|A(t_0)\|)^{n-1}\\ &\leq \|A-B\|\,(\|A\|+\|B\|)^{n-1}.\end{aligned}$$

Hence,

$$|b-a| \leq 4\cdot 2^{-1/n}\,(\|A\|+\|B\|)^{1-1/n}\,\|A-B\|^{1/n},$$

and this proves the theorem. ■

G. Krause has shown that the factor 4 in (39.2) can be replaced by 3.08. A sharp bound will be of much interest.

Notes and references

The argument involving Chebyshev polynomials, in Sections 38 and 39, seems to have been first used by A. Schönhage [S], and then rediscovered by D. Phillips [P] who proved an inequality weaker than (39.2) with a factor 8 instead of 4. The arguments of Phillips were simplified and improved by R. Bhatia, L. Elsner and G. Krause [BEK] who proved (38.4), (38.7) and (39.2). The further improvement mentioned in Sections 38 and 39 is given in [Kr].

12 Arbitrary perturbations of constrained matrices

The theorem in Section 24.1 has been improved as well as extended in scope. This is presented in the next section.

§40 Arbitrary perturbations of normal matrices

Let A be any matrix and let A_D be the diagonal part of A, and A_L, A_U its parts below and above the diagonal. Thus

$$A = A_L + A_D + A_U.$$

Proposition 40.1 Let A be an $n \times n$ normal matrix. Then

$$\|A_L\|_F \le \sqrt{n-1}\, \|A_U\|_F, \quad \|A_U\|_F \le \sqrt{n-1}\, \|A_L\|_F.$$

Proof If A is normal, then the Euclidean norms of its jth row and jth column are equal. So, if A is partitioned as

$$A = \begin{bmatrix} V & W \\ X & Y \end{bmatrix}$$

where V is a $k \times k$ matrix, $1 \le k \le n-1$, then $\|W\|_F^2 = \|X\|_F^2$. Summing up all these equalities over $k = 1, 2, \ldots, n-1$, we get

$$\sum_{j=1}^{n-1} \sum_{\ell > j} (\ell - j) |a_{j\ell}|^2 = \sum_{\ell=1}^{n-1} \sum_{j > \ell} (j - \ell) |a_{j\ell}|^2.$$

This is used at the third step in the following calculation.

$$\begin{aligned}\|A_U\|_F^2 &= \sum_{j=1}^{n-1}\sum_{\ell>j}|a_{j\ell}|^2 \le \sum_{j=1}^{n-1}\sum_{\ell>j}(\ell-j)|a_{j\ell}|^2 \\ &= \sum_{\ell=1}^{n-1}\sum_{j>\ell}(j-\ell)|a_{j\ell}|^2 \\ &\le (n-1)\sum_{\ell=1}^{n-1}\sum_{j>\ell}|a_{j\ell}|^2 = (n-1)\|A_L\|_F^2.\end{aligned}$$

This proves one of the assertions. By symmetry, the other is also true. ■

If A is the $n \times n$ unitary matrix

$$A = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \end{bmatrix},$$

then $\|A_U\|_F = \sqrt{n-1}\,\|A_L\|_F$. This shows that the inequalities in the proposition above are best possible.

Theorem 40.2 Let A and B be $n \times n$ matrices and suppose A is normal. Then

$$\|(\operatorname{Eig} A,\ \operatorname{Eig} B)\|_F \le \sqrt{n}\,\|A - B\|_F. \tag{40.1}$$

Proof We can apply a unitary similarity to A and B, and assume that B is upper triangular. In this basis let

$$A = A_L + A_D + A_U, \qquad B = B_D + B_U.$$

The matrix B_D is normal and has the same eigenvalues as those of B. Therefore, from Theorem 15.1 we get

$$\|(\operatorname{Eig} A,\ \operatorname{Eig} B)\|_F \le \|A - B_D\|_F.$$

Since $B_L = 0$ we have

$$A - B_D = (A-B)_D + (A-B)_L + A_U.$$

So

$$\begin{aligned}\|A - B_D\|_F^2 &= \|(A-B)_D\|_F^2 + \|(A-B)_L\|_F^2 + \|A_U\|_F^2 \\ &\leq \|(A-B)_D\|_F^2 + \|(A-B)_L\|_F^2 + (n-1)\|A_L\|_F^2 \\ &= \|(A-B)_D\|_F^2 + n\ \|(A-B)_L\|_F^2 \\ &\leq n\ \|A-B\|_F^2.\end{aligned}$$

This proves the theorem. ■

Remark 40.3 If A is Hermitian, then $\|A_U\|_F = \|A_L\|_F$. Then the argument above leads to a stronger inequality

$$\|(\mathrm{Eig}\,A,\ \mathrm{Eig}\,B)\|_F \leq \sqrt{2}\ \|A-B\|_F. \tag{40.2}$$

Remark 40.4 The Schatten p-norms and the Frobenius norm $(p=2)$ are related by the inequalitites

$$\begin{aligned}\|A\|_F &\leq \|A\|_p \leq n^{1/p-1/2}\|A\|_F, \quad 1 \leq p \leq 2, \\ \|A\|_p &\leq \|A\|_F \leq n^{1/2-1/p}\|A\|_p, \quad 2 \leq p \leq \infty.\end{aligned}$$

So, from (40.1) we obtain two families of inequalities (valid if A is normal and B arbitrary):

$$\|(\mathrm{Eig}\,A,\ \mathrm{Eig}\,B)\|_p \leq n^{1/p}\ \|A-B\|_p, \quad 1 \leq p \leq 2, \tag{40.3}$$

$$\|(\mathrm{Eig}\,A,\ \mathrm{Eig}\,B)\|_p \leq n^{1-1/p}\ \|A-B\|_p, \quad 2 \leq p \leq \infty. \tag{40.4}$$

The case $p=\infty$ is the operator norm, and we have

$$\|(\mathrm{Eig}\,A,\ \mathrm{Eig}\,B)\| \leq n\ \|A-B\|. \tag{40.5}$$

This is an improvement on (24.1) where we had $(2n-1)$ instead of n.

The inequalities (40.3) are best possible. This can be seen by choosing A to be the unitary matrix displayed before Theorem 40.2, and B to be the matrix obtained by replacing the bottom left entry of A by zero. The same example shows that the bound (40.2) is attained for 2×2 matrices.

Remark 40.5 When A is Hermitian and B normal, special results hold. In this case

$$\|(\mathrm{Eig}\,A,\ \mathrm{Eig}\,B)\|_p \ \leq\ 2^{2/p-1}\ \|A-B\|_p, \quad 1 \leq p \leq 2, \tag{40.6}$$

$$\|(\mathrm{Eig}\,A,\ \mathrm{Eig}\,B)\|_p \ \leq\ 2^{1/2-1/p}\ \|A-B\|_p, \quad 2 \leq p \leq \infty. \tag{40.7}$$

This was observed by R. Bhatia and L. Elsner. The case $p=\infty$ of (40.7) says

$$\|(\mathrm{Eig}\,A,\ \mathrm{Eig}\,B)\| \ \leq\ \sqrt{2}\ \|A-B\|,$$

and this is the inequality (23.9) noted earlier.

Notes and references

Results in Section 40 are due to J.-G. Sun [S2], except for those in Remark 40.5 which can be found in [BE].

13 Related Topics

§41 Operators on infinite-dimensional spaces

Let $\mathcal{H}$ be a complex, separable, and infinite-dimensional Hilbert space, and $\mathcal{B}(\mathcal{H})$ the space of bounded linear operators on $\mathcal{H}$. For every A in $\mathcal{B}(\mathcal{H})$ its spectrum $\operatorname{Spec} A$ is a compact subset of the complex plane. The map $A \mapsto \operatorname{Spec} A$, in general, is discontinuous: a small change in the operator can cause a big change in the spectrum. However, on the special class of normal operators, the spectrum is continuous. More precisely, if $h(E, F)$ stands for the Hausdorff distance between two compact subsets of the plane, then

$$h(\operatorname{Spec} A,\ \operatorname{Spec} B) \ \leq\ \|A - B\|,$$

whenever A and B are normal. The proof is a slight modification of the finite-dimensional case as given for Theorem 12.4.

Quantitative estimates for the distance between spectra are far more intricate and difficult to obtain. In this section we briefly outline results that are parallel to the ones we have given for the finite-dimensional case.

The simplest case to consider is that of compact self-adjoint operators A and B. All nonzero points in $\operatorname{Spec} A$ are isolated eigenvalues with finite multiplicities. The point zero is either an eigenvalue of infinite multiplicity or a limit point of the nonzero eigenvalues.

In the finite-dimensional case we arranged the eigenvalues of A and B in decreasing order, and then paired them as in Theorems 8.1 and 9.7. The presence of zero amidst the spectra creates problems in the infinite-dimensional case and there are different ways of getting around them. Let $\{\alpha_j : j = 1, 2, \ldots\}$ and $\{\alpha_{-j} : j = 1, 2, \ldots\}$ be two infinite sequences with the following properties

(i) $\alpha_1 \geq \alpha_2 \geq \cdots \geq 0$,
$\alpha_{-1} \leq \alpha_{-2} \leq \cdots \leq 0$;

(ii) all nonzero eigenvalues of A are included in these sequences with their proper multiplicities;

(iii) if A has infinitely many positive eigenvalues, then these make up the entire sequence $\{\alpha_j : j = 1, 2, \ldots\}$; but if A has finitely many positive eigenvalues, then in addition to them this sequence contains infinitely many zeros;

(iv) similarly the sequence $\{\alpha_{-j} : j = 1, 2, \ldots\}$ contains only the negative eigenvalues of A if there are infinitely many of them, and if there are only a finite number of these, then it contains in addition to them infinitely many zeros.

The collection $\{\alpha_j : j = \pm 1, \pm 2, \ldots\}$ is called an *enumeration* of the eigenvalues of A.

Let A and B be compact operators and suppose $A - B$ is in the Schatten class C_p for some $1 \leq p \leq \infty$. Let $\{\alpha_j : j \pm 1, \pm 2, \ldots\}$ and $\{\beta_j : j = \pm 1, \pm 2, \ldots\}$ be enumerations of the eigenvalues of A and B, respectively. Then

$$\left(\sum_{j=1}^{\infty} |\alpha_j - \beta_j|^p + \sum_{j=1}^{\infty} |\alpha_{-j} - \beta_{-j}|^p \right)^{1/p} \leq \|A - B\|_p. \tag{41.1}$$

A similar statement is true for any unitarily invariant norm $\|\cdot\|_\Phi$.

This theorem, due to A. S. Markus, is analogous to Theorem 9.7. There is another kind of result in which it is not necessary to add a spurious zero to the genuine eigenvalues of A and B, but the information about the pairing of eigenvalues is lost. This says that if A and B are Hermitian operators in C_p for some $1 < p < \infty$, and $\{\alpha_j\}$ and $\{\beta_j\}$ are the eigenvalues of A and B where each eigenvalue is counted with its proper multiplicity, then there exists a permutation σ of $\mathbb{N}$ such that

$$\left(\sum_j \left|\alpha_j - \beta_{\sigma(j)}\right|^p \right)^{1/p} \leq \|A - B\|_p. \tag{41.2}$$

J. A. Cochran and E. W. Hinds proved that for $p = 2$, the inequality (41.2) is true, more generally, for normal operators A and B. This is the Hoffman-Wielandt inequality extended to compact (Hilbert-Schmidt) operators.

Next we consider the case when A and B are any two bounded normal operators whose difference $A - B$ is compact. In this case good analogues of perturbation bounds of Chapters 3 and 4 have been established.

The spectrum of a normal operator A is a disjoint union of two sets, the *discrete spectrum* consisting of those isolated points of the spectrum which are eigenvalues of finite multiplicity, and the *essential spectrum.* If P is the spectral measure corresponding to A, then a point λ is in the essential spectrum if and only if for every neighbourhood E of λ the projection $P(E)$ is infinite-dimensional. The essential spectrum is a closed subset of the plane. There is a theorem of H. Weyl that says that if $A - B$ is compact, then the essential spectra of A and B coincide. Thus taking the discrete spectrum as the kind of spectrum familiar from linear algebra, we may seek for it perturbation bounds resembling the matrix case. Here a problem arises. A compact perturbation could completely wipe off the discrete spectrum; i.e., the operator A could have a nonempty discrete spectrum while that of B could be empty even though $A - B$ is compact. To get around this difficulty T. Kato introduced an interesting idea.

An *extended enumeration* of the discrete eigenvalues of A is any sequence $\{\alpha_j\}$ that includes all points of the discrete spectrum of A each counted as often as its multiplicity as an eigenvalue, and that in addition may include some boundary points of the essential spectrum of A. (Not all boundary points of the essential spectrum are required to be present in $\{\alpha_j\}$, and those that are may be repeated arbitrarily often.)

In 1987, Kato proved the following

Theorem 41.1 Let A and B be self-adjoint operators on $\mathcal{H}$ such that $A - B$ is a compact operator, and is in the class C_p. Then there exist extended enumerations $\{\alpha_j\}$ and $\{\beta_j\}$ of discrete eigenvalues of A and B such that

$$\left(\sum_{j=1}^{\infty} |\alpha_j - \beta_j|^p\right)^{1/p} \leq \|A - B\|_p. \tag{41.3}$$

A similar statement is true for all unitarily invariant norms $\|\cdot\|_\Phi$.

This theorem includes in it the result of Markus as a special case. (When A is compact the only possible point in its essential spectrum is 0.)

Kato's paper was soon followed by one by R. Bhatia and K. B. Sinha who proved a similar extension of Theorem 14.7 for unitary operators A and B whose difference is compact. In the finite-dimensional case, these results for Hermitian and unitary operators and a few more, can be derived as corollaries of Theorem 14.3. An extension of this general theorem to infinite dimensions was published by R. Bhatia and C. Davis in 1999.

Theorem 41.2 Let A_0 and A_1 be normal operators whose difference $A_0 - A_1$ is in the class C_Φ corresponding to a unitarily-invariant norm $\|\cdot\|_\Phi$. Let $A(t)$, $0 \le t \le 1$, be a piecewise C^1 curve such that

(i) $A(t)$ is normal for all t,

(ii) $A(0) = A_0$, $A(1) = A_1$,

(iii) $A(t) - A_0$ is in C_Φ for all t.

Then there exist extended enumerations $\{\lambda_j(0)\}$ and $\{\lambda_j(1)\}$ of discrete eigenvalues of A_0 and A_1, respectively, for which

$$\Phi\left(\{\lambda_j(0) - \lambda_j(1)\}\right) \le \int_0^1 \|A'(t)\|_\Phi \, dt.$$

This is a master theorem, from which we can derive as in Chapter 4, analogues of Theorem 13.6 (A and B unitary), Theorem 14.5 (A, B and $A - B$ normal), and Theorem 14.7 (A and B unitary). Two theorems in which A and B are arbitrary normal operators do need separate proofs. These have been proved by R. Bhatia and C. Davis (analogue of Theorem 16.6) and by L. Elsner and S. Friedland (analogue of Theorem 15.1). The formulations involve extended enumerations as in the other results stated above and the proofs need suitable extensions of the marriage lemma.

§42 Joint spectra of commuting matrices

In 1991 A. Pryde initiated a program of extending the perturbation bounds in this book to commuting tuples of matrices. We briefly indicate the nature of the problem and the results obtained.

Let $\mathbf{A} = \left(A^{(1)}, \ldots, A^{(m)}\right)$ be an m-tuple of pairwise commuting operators on an n-dimensional Hilbert space $\mathcal{H}$. A point $\underline{\lambda} = \left(\lambda^{(1)}, \ldots, \lambda^{(m)}\right)$ in $\mathbb{C}^m$ is called a *joint eigenvalue* of $\mathbf{A}$ if there exists a vector x in $\mathcal{H}$ such that

$$A^{(j)}x = \lambda^{(j)}x \quad 1 \leq j \leq m.$$

The vector x is called a *joint eigenvector* of $\mathbf{A}$ corresponding to the joint eigenvalue $\underline{\lambda}$.

By the Schur triangularization theorem there exists an orthonormal basis of $\mathcal{H}$ in which the matrix of each operator $A^{(j)}$ is upper triangular:

$$A^{(j)} = \begin{bmatrix} \lambda_1^{(j)} & * & * & \ldots & * \\ 0 & \lambda_2^{(j)} & * & \ldots & * \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \ldots & \ldots & \lambda_n^{(j)} \end{bmatrix}, \quad 1 \leq j \leq m.$$

This is called a Schur basis for $\mathbf{A}$. The joint eigenvalues of $\mathbf{A}$ are the n points $\underline{\lambda}_k = \left(\lambda_k^{(1)}, \lambda_k^{(2)}, \ldots, \lambda_k^{(m)}\right)$, $1 \leq k \leq n$. If the operators $A^{(j)}$ are normal, then in a Schur basis, they are all diagonal.

We can view the m-tuple $\mathbf{A}$ as a column vector

$$\mathbf{A} = \begin{bmatrix} A^{(1)} \\ A^{(2)} \\ \vdots \\ A^{(m)} \end{bmatrix}$$

and think of it as an operator from $\mathcal{H}$ to the space $\mathcal{H} \oplus \mathcal{H} \oplus \cdots \oplus \mathcal{H}$ (m copies). The Frobenius norm of $\mathbf{A}$ is defined as

$$\|\mathbf{A}\|_F = (\operatorname{tr} \mathbf{A}^*\mathbf{A})^{1/2} = \left(\sum_{j=1}^{m} \|A^{(j)}\|_F^2\right)^{1/2}. \tag{42.1}$$

The following theorem is an analogue of the Hoffman-Wielandt inequality.

Theorem 42.1 Let $\mathbf{A} = \left(A^{(1)}, \ldots, A^{(m)}\right)$ and $\mathbf{B} = \left(B^{(1)}, \ldots, B^{(m)}\right)$ be two m-tuples of pairwise commuting normal operators on an n-dimensional Hilbert space $\mathcal{H}$. Let $\underline{\alpha}_k$ and $\underline{\beta}_k$, $1 \leq k \leq n$, be the joint eigenvalues of

A and **B**, respectively. Then there exists a permutation σ on n indices such that

$$\sum_{k=1}^{n} \|\underline{\alpha}_k - \underline{\beta}_{\sigma(k)}\|_{\mathbb{C}^m}^2 \le \|\mathbf{A} - \mathbf{B}\|_F^2. \qquad (42.2)$$

The norm occuring in the expression on the left is the Euclidean vector norm on $\mathbb{C}^m$. The inequality can be written in another way as

$$\sum_{k=1}^{n}\sum_{j=1}^{m} \left|\alpha_k^{(j)} - \beta_{\sigma(k)}^{(j)}\right|^2 \le \sum_{j=1}^{m} \left\|A^{(j)} - B^{(j)}\right\|_F^2. \qquad (42.3)$$

when $m = 1$, this is the Hoffman-Wielandt inequality. The noteworthy feature of (42.3) is that the *same* permutation σ does the job for each of the components.

This inequality was first proved by R. Bhatia and T. Bhattacharyya. Soon afterwards, L. Elsner gave a much simpler proof. In the special case $m = 1$ this becomes a very economical proof of the original Hoffman - Wielandt theorem. Here is how this argument goes.

Choose unitary operators U and V such that for all $1 \le j \le m$, the operators $UA^{(j)}U^*$ and $VB^{(j)}V^*$ are diagonal; i.e.,

$$\begin{aligned} UA^{(j)}U^* &= \Lambda^{(j)} = \operatorname{diag}\left(\alpha_1^{(j)}, \alpha_2^{(j)}, \ldots, \alpha_n^{(j)}\right) \\ VB^{(j)}V^* &= \Gamma^{(j)} = \operatorname{diag}\left(\beta_1^{(j)}, \beta_2^{(j)}, \ldots, \beta_n^{(j)}\right). \end{aligned}$$

Then

$$\begin{aligned} \sum_{j=1}^{m} \|A^{(j)} - B^{(j)}\|_F^2 &= \sum_{j=1}^{m} \|U^*\Lambda^{(j)}U - V^*\Gamma^{(j)}V\|_F^2 \\ &= \sum_{j=1}^{m} \|\Lambda^{(j)}UV^* - UV^*\Gamma^{(j)}\|_F^2 \\ &= \sum_{j=1}^{m} \|\Lambda^{(j)}W - W\Gamma^{(j)}\|_F^2 \\ &= \sum_{j=1}^{m}\sum_{k,\ell=1}^{n} \left|\alpha_k^{(j)} - \beta_\ell^{(j)}\right|^2 |w_{k\ell}|^2, \qquad (42.4) \end{aligned}$$

where $W = UV^*$ is an $n \times n$ unitary matrix, and its entries are $w_{k\ell}$. The matrix whose entries are $|w_{k\ell}|^2$ is doubly stochastic. Let S be any

$n \times n$ doubly stochastic matrix and let

$$f(S) = \sum_{j=1}^{m} \sum_{k,\ell=1}^{n} \left| \alpha_k^{(j)} - \beta_\ell^{(j)} \right|^2 s_{k\ell}.$$

Then f is an affine linear functional on the set Ω_n consisting of $n \times n$ doubly stochastic matrices. Hence the minimum value of f is attained at one of the extreme points of Ω_n. By Birkhoff's theorem such a point is a permutation matrix P. Thus the expression in (42.4) is greater than or equal to

$$\sum_{j=1}^{m} \sum_{k,\ell=1}^{n} \left| \alpha_k^{(j)} - \beta_\ell^{(j)} \right|^2 p_{k\ell}.$$

The matrix P is associated with a permutation σ on $\{1, 2, \ldots, n\}$, and this expression, in turn, is equal to

$$\sum_{j=1}^{m} \sum_{k,\ell=1}^{n} \left| \alpha_k^{(j)} - \beta_{\sigma(k)}^{(j)} \right|^2.$$

This proves the inequality (42.3).

Instead of the Frobenius norm (42.1) we could consider

$$\|\mathbf{A}\| = \|\mathbf{A}^*\mathbf{A}\|^{1/2} = \left\| \sum_{j=1}^{m} A^{(j)*} A^{(j)} \right\|^{1/2}, \tag{42.5}$$

and expect analogues of Theorem 8.1 and 16.6. No good results of this kind have been proved.

An upper bound for $\|\mathbf{A}-\mathbf{B}\|$ analogous to (33.6) has been established. If $\mathbf{A}$ and $\mathbf{B}$ are two m-tuples of commuting normal operators, then

$$\|\mathbf{A} - \mathbf{B}\| \leq \sqrt{2} \max_{i,j} \|\underline{\alpha}_i - \underline{\beta}_j\|_{\mathbb{C}^m}. \tag{42.6}$$

Analogues of some other theorems like those of Bauer-Fike, and Henrici have been proved. The distance between the tuples $\mathbf{A}$ and $\mathbf{B}$ is defined via *Clifford operators* associated with them. We do not state these theorems here. More theorems, and better ones, can perhaps be proved in this context.

§43 Relative perturbation theory

All the bounds studied in this book have two features. We restricted ourselves to *additive* perturbations: the matrix A is perturbed to $B = A + E$ and E is regarded as the error or the perturbation. We obtained *absolute* error bounds: the quantities for which we found bounds involved the absolute errors $|\alpha_j - \beta_j|$, and all of these at the same time.

There are other kinds of perturbations and perturbation bounds. We illustrate the problems with an old theorem of A. Ostrowski.

Let A be a Hermitian matrix, and let X be any nonsingular matrix. Then $B = X^*AX$ may be thought of as a *multiplicative perturbation* of A. How are the eigenvalues of A and B related? Sylvester's Law tells us that the inertias of A and B are equal. Finer information is provided by a theorem of Ostrowski: there exist positive real numbers t_k, such that $\lambda_{[n]}(X^*X) \leq t_k \leq \lambda_{[1]}(X^*X)$ and

$$\lambda_{[k]}(B) = t_k \lambda_{[k]}(A), \quad 1 \leq k \leq n. \tag{43.1}$$

From this we see that

$$\max_{1 \leq k \leq n} \left| \frac{\lambda_{[k]}(B) - \lambda_{[k]}(A)}{\lambda_{[k]}(A)} \right| \leq \|I - X^*X\|, \tag{43.2}$$

provided $\lambda_{[k]}(A) \neq 0$. The quantities on the left now involve not the absolute errors $|\alpha_k - \beta_k|$ but the relative errors $|\alpha_k - \beta_k|/|\alpha_k|$.

The subject of relative perturbation bounds has seen much activity in the last fifteen years, and very interesting theorems of importance in numerical analysis have been proved by several authors. The reader can find a convenient summary and references to some of the important papers in the two survey articles by I. C. F. Ipsen and R.-C. Li.

Notes and references

The inequality (41.1) is given in [M] and (41.2) in [BSe]. The paper [CW] contains the extension of the Hoffman-Wielandt inequality to normal Hilbert-Schmidt operators, and [EF] to normal operators whose difference is Hilbert-Schmidt. The paper by T. Kato [Ka] stimulated the subsequent work of [BSi] and [BD]. A recent paper [KMM] contains a study of other infinite-dimensional problems related to the material in this book.

The stimulus for the work in Section 42 came from [Pr1] and [Pr2]. This led to the work in [BB1], [BB2] and [E].

Much work in relative perturbation theory has been done in recent years by several authors. We do not attempt to list the important papers. Instead we refer the reader to the survey [I], and a recent handbook article [Li] where an up-to-date summary of both the absolute and the relative perturbation theory and a comprehensive bibliography are included.

Bibliography

[AB] T. Ando and R. Bhatia, *Eigenvalue inequalities associated with the Cartesian decomposition,* Linear and Multilinear Algbera, 22(1987) 133-147. (Cited on pp. 139, 159.)

[BGLS] H. Bauschke, O. Güler, A. S. Lewis, and H. S. Sendov, *Hyperbolic polynomials and convex analysis,* Canad. J. Math., 53 (2001) 470-488. (Cited on p. 151.)

[B1] R. Bhatia, *Matrix Analysis,* Springer, 1997. (Cited on pp. 139, 151.)

[B2] R. Bhatia, *Some inequalities for norm ideals,* Commun. Math. Phys. 111 (1987) 33-39. (Cited on p. 139.)

[B3] R. Bhatia, *Linear algebra to quantum cohomology: The story of Alfred Horn's inequalities,* Amer. Math. Monthly, 108 (2001) 289-318. (Cited on p. 151.)

[BB1] R. Bhatia and T. Bhattacharyya, *A generalization of the Hoffman-Wielandt theorem,* Linear Algebra Appl., 179 (1993) 11-17. (Cited on p. 184.)

[BB2] R. Bhatia and T. Bhattacharyya, *A Henrici theorem for joint spectra of commuting matrices,* Proc. Amer. Math. Soc., 118 (1993) 5-14. (Cited on p. 184.)

[BD] R. Bhatia and C. Davis, *Perturbation of extended enumerations of eigenvalues,* Acta Sci. Math. (Szeged), 65 (1999) 277-286. (Cited on p. 183.)

[BDKi] R. Bhatia, C. Davis, and F. Kittaneh, *Some inequalities for commutators and an application to spectral variation,* Aequationes Math., 41 (1991) 70-78. (Cited on p. 166.)

[BDKo] R. Bhatia, C. Davis, and P. Koosis, *An extremal problem in Fourier analysis with applications to operator theory,* J. Functional Anal., 82 (1989) 138-150. (Cited on p. 159.)

[BE] R. Bhatia and L. Elsner, *The Hoffman-Wielandt inequality in infinite dimensions,* Proc. Indian Acad. Sci. (Math. Sci.), 104 (1994), 483-494. (Cited on p. 175.)

[BEK] R. Bhatia, L. Elsner, and G. Krause, *Bounds for the variation of the roots of a polynomial and the eigenvalues of a matrix,* Linear Algebra Appl., 142 (1990) 195-209. (Cited on p. 171.)

[BES] R. Bhatia, L. Elsner, and P. Šemrl, *Distance between commuting tuples of normal operators,* Arch. Math., 71 (1998) 229-232. (Cited on p. 159.)

[BH1] R. Bhatia and J. A. R. Holbrook, *Unitary invariance and spectral variation,* Linear Algebra Appl., 95 (1987) 43-68.

[BH2] R. Bhatia and J. A. R. Holbrook, *On the Clarkson-McCarthy inequalities,* Math. Ann., 281 (1988) 7-12. (Cited on p. 139.)

[BH3] R. Bhatia and J. A. R. Holbrook, *A softer, stronger Lidskii theorem,* Proc. Indian Acad. Sci. (Math. Sci.) 99 (1989) 75-83. (Cited on p. 139.)

[BK] R. Bhatia and F. Kittaneh, *Cartesian decompositions and Schatten norms,* Linear Algebra Appl., 318 (2000) 109-116. (Cited on p. 160.)

[BKL] R. Bhatia, F. Kittaneh, and R.-C. Li, *Some inequalities for commutators and an application to spectral variation II,* Linear and Multilinear Algebra, 43 (1997) 207-219. (Cited on pp. 139, 166.)

[BR] R. Bhatia and P. Rosenthal, *How and why to solve the operator equation* $AX-XB=Y$, Bull. London Math. Soc., 29 (1997) 1-21. (Cited on p. 166.)

[BSe] R. Bhatia and P. Šemrl, *Distance between Hermitian operators in Schatten classes,* Proc. Edinburgh Math. Soc., 39 (1996) 377-380. (Cited on p. 183.)

[BSi] R. Bhatia and K. B. Sinha, *A unitary analogue of Kato's theorem on variation of discrete spectra,* Letters Math. Phys., 15 (1998) 201-204. (Cited on p. 183.)

[BZ1] R. Bhatia and X. Zhan, *Compact operators whose real and imaginary parts are positive,* Proc. Amer. Math. Soc., 129 (2001) 2277-2281. (Cited on p. 160.)

[BZ2] R. Bhatia and X. Zhan, *Norm inequalities for operators with positive real part,* J. Operator Theory, 50 (2003) 67-76. (Cited on p. 160.)

[CW] J. A. Cochran and E. W. Hinds, *Improved error bounds for the eigenvalues of certain normal operators,* SIAM J. Numer. Anal., 9 (1972) 446-453. (Cited on p. 183.)

[D] J. Demmel, *Applied Numerical Linear Algebra,* SIAM, 1997. (Cited on p. 139.)

[DK] C. Davis and W. M. Kahan, *The rotation of eigenvectors by a perturbation III,* SIAM J. Numer. Anal., 7 (1968) 1-46. (Cited on p. 139.)

[E] L. Elsner, *A note on the Hoffman-Wielandt theorem,* Linear Algebra Appl., 182 (1993) 235-237. (Cited on p. 184.)

[EF] L. Elsner and S. Friedland, *Variation of the discrete eigenvalues of normal operators,* Proc. Amer. Math. Soc., 123 (1995) 2511-2517. (Cited on p. 183.)

[EH] L. Elsner and C. He, *Perturbation and interlace theorems for the unitary eigenvalue problem,* Linear Algebra Appl., 188/189 (1993) 207-229.

[Fa] F. O. Farid, *The spectral variation of two matrices with spectra on two intersecting lines,* Linear Algebra Appl., 177 (1992) 251-273.

[F] W. Fulton, *Eigenvalues, invariant factors, highest weights, and Schubert calculus,* Bull. Amer. Math. Soc., 37 (2000) 209-249. (Cited on p. 151.)

[GR] L. Gurvits and L. Rodman, *On matrix polynomials with real roots,* SIAM J. Matrix Anal. Appl., 26 (2005) 758-764. (Cited on p. 151.)

[GV] G. H. Golub and C. F. Van Loan, *Matrix Computations,* 3rd ed., Johns Hopkins University Press, 1996. (Cited on p. 139.)

[HN] F. Hiai and Y. Nakamura, *Majorization for generalized s-numbers in semifinite von Neumann algebras,* Math. Z., 195 (1987) 17-27. (Cited on p. 139.)

[Hi] N. J. Higham, *Accuracy and Stability of Numerical Algorithms,* 2nd ed., SIAM, 2002. (Cited on p. 139.)

[Ho] J. A. R. Holbrook, *Spectral variation of normal matrices,* Linear Algbera Appl., 174 (1992) 131-144 (Cited on p. 159.)

[HB] L. Hörmander and B. Bernhardsson, *An extension of Bohr's inequality,* in Boundary Value Problems of Partial Differential Equations and Applications, J.-L. Lions and C. Baiocchi eds., Masson, 1993. (Cited on p. 159.)

[HM] R. A. Horn and R. Mathias, *Cauchy-Schwarz inequalities associated with positive semidefinite matrices,* Linear Algebra Appl., 142 (1990) 63 - 82. (Cited on p. 139.)

[HV] J. W. Helton and V. Vinnikov, *Linear matrix inequality representation of sets,* Technical report, University of California, San Diego, 2002. (Cited on p. 151.)

[HY] J.-B. Hiriart-Urruty and D. Ye, *Sensitivity analysis of all eigenvalues of a symmetric matrix,* Numer. Math., 70 (1992) 45-72. (Cited on p. 151.)

[I] I. C. F. Ipsen, *Relative perturbation results for matrix eigenvalues and singular values,* Acta Numer., 7 (1998) 151-201. (Cited on p. 184.)

[K] W. Kahan, *Inclusion theorems for clusters of eigenvalues of Hermitian matrices,* Technical Report, Computer Science Department, University of Toronto, 1967. (Cited on p. 166.)

[Ka] T. Kato, *Variation of discrete spectra,* Commun. Math. Phys., 111 (1987) 501-504. (Cited on p. 183.)

[Kl] A. A. Klyachko, *Stable bundles, representation theory and Hermitian operators,* Selecta Math. 4 (1998) 419-445. (Cited on p. 151.)

[KMM] V. Kostrykin, K. A. Makarov, and A. K. Motovilov, *Perturbation of spectra and spectral subspaces,* Trans. Amer. Math. Soc., 359 (2007) 77-89. (Cited on p. 183.)

[KT] A. Knutson and T. Tao, *The honeycomb model of $GL_n(\mathbb{C})$ tensor products I: Proof of the saturation conjecture,* J. Amer. Math. Soc., 12 (1999) 1055-1090. (Cited on p. 151.)

[Kr] G. Krause, *Bounds for the variation of matrix eigenvalues and polynomial roots,* Linear Algebra Appl., 208/209 (1994) 73-82. (Cited on p. 171.)

[Le] A. S. Lewis, *Lidskii's theorem via nonsmooth analysis,* SIAM J. Matrix Anal. Appl., 21 (1999) 379-381. (Cited on p. 151.)

[LPR] A. S. Lewis, P. A. Parrilo, and M. V. Ramana, *The Lax conjecture is true,* Proc. Amer. Math. Soc., 133 (2005) 2495-2499. (Cited on p. 151.)

[LL] C.-K. Li and R.-C. Li, *A note on eigenvalues of perturbed Hermitian matrices,* Linear Algebra Appl., 395 (2005) 183-190. (Cited on p. 151.)

[LM1] C.-K. Li and R. Mathias, *Generalizations of Ky Fan's dominance theorem,* SIAM J. Matrix Anal. Appl., 19 (1998) 99-106. (Cited on p. 139.)

[LM2] C.-K. Li and R. Mathias, *The Lidskii-Mirsky-Wielandt theorem—additive and multiplicative versions,* Numer. Math., 81 (1999) 377-413. (Cited on p. 151.)

[LTS] C.-K Li, T.-Y. Tam and N.-K. Tsing, *The generalized spectral radius, numerical radius and spectral norm,* Linear and Multilinear Algebra, 16 (1984) 215-237. (Cited on p. 139.)

[Li] R.-C. Li, *Matrix Perturbation Theory,* in Handbook of Linear Algebra, L. Hogben ed., Chapman and Hall/CRC, 2006. (Cited on p. 184.)

[Lo] K. Löwner, *Über monotone Matrixfunctionen,* Math. Z., 38 (1934) 177-216. (Cited on p. 150.)

[Lu] T.-X. Lu, *Perturbation bounds for eigenvalues of symmetrizable matrices,* Numerical Mathematics: a Journal of Chinese Universities, 16 (1994) 177-185 (in Chinese). (Cited on p. 166.)

[M] A. S. Markus, *The eigen and singular values of the sum and product of linear operators,* Russian Math. Surveys, 19 (1964) 92-120. (Cited on p. 183.)

[P] D. Phillips, *Improving spectral variation bounds with Chebyshev polynomials,* Linear Algebra Appl., 133 (1990) 165-173. (Cited on p. 171.)

[Pr1] A. Pryde, *A Bauer-Fike theorem for the joint spectrum of commuting matrices,* Linear Algebra Appl., 173 (1992) 221-230. (Cited on p. 184.)

[Pr2] A. Pryde, *Optimal matching of joint eigenvalues of normal matrices,* Monash University Analysis Paper 74, March, 1991. (Cited on p. 184.)

[S] A. Schönhage, *Quasi-GCD computations,* J. Complexity, 1 (1985) 118-137. (Cited on p. 171.)

[SS] G. W. Stewart and J.-G. Sun, *Matrix Perturbation Theory,* Academic Press, 1990. (Cited on p. 139.)

[S1] J.-G. Sun, *On the perturbation of the eigenvalues of a normal matrix,* Math. Numer. Sinica, 6 (1984) 334-336. (Cited on p. 166.)

[S2] J.-G. Sun, *On the variation of the spectrum of a normal matrix,* Linear Algebra Appl., 246 (1996) 215-222. (Cited on p. 175.)

[W] H. Wielandt, *Mathematical Works,* Vol. 2, B. Huppert and H. Schneider eds., Walter de Gruyter, 1996. (Cited on p. 151.)

[V] V. Vinnikov, *Self-adjoint determinantal representations of real plane curves,* Math. Ann., 296 (1993) 453-479. (Cited on p. 151.)

[Z] X. Zhan, *Norm inequalities for Cartesian decomposition,* Linear Algebra Appl., 286 (1999) 297-301. (Cited on p. 160.)

Errata

Page 20, last line :	replace the second "min" by "max"
Page 33, line $3_\uparrow$:	replace "Naumann" by "Neumann"
Page 36, line $11_\downarrow$:	replace "a vector" by "a unit vector"
Page 91, line $7_\uparrow$:	replace " θ_k " by " α_k "
Page 120, line $6_\uparrow$:	replace " $n-1$ " by " n "

Perturbation Bounds for Matrix Eigenvalues contains a unified exposition of spectral variation inequalities for matrices. The text provides a complete and self-contained collection of bounds for the distance between the eigenvalues of two matrices, which could be arbitrary or restricted to special classes. The book's emphasis on sharp estimates, general principles, elegant methods, and powerful techniques makes it a good reference for researchers and students.

For the SIAM Classics edition, the author has added over 60 pages of new material, which includes recent results and discusses the important advances made in the theory, results, and proof techniques of spectral variation problems in the two decades since the book's original publication.

This updated edition is appropriate for use as a research reference for physicists, engineers, computer scientists, and mathematicians interested in operator theory, linear algebra, and numerical analysis. The text is also suitable for a graduate course in linear algebra or functional analysis.

Rajendra Bhatia is a Professor at the Indian Statistical Institute and has held visiting positions at several universities around the world. He is a J. C. Bose National Fellow and a fellow of the Indian Academy of Sciences and of the Indian National Science Academy. Professor Bhatia received the Bhatnagar Prize for Science and Technology in 1995. He is a Senior Editor of *Linear Algebra and Its Applications* and a past editor of *SIAM Journal on Matrix Analysis and Its Applications*.

For more information about SIAM books, journals, conferences, memberships, or activities, contact:

siam®

Society for Industrial and Applied Mathematics
3600 University City Science Center
Philadelphia, PA 19104-2688 USA
+1-215-382-9800 • Fax: +1-215-386-7999
siam@siam.org • *www.siam.org*

ISBN 978-0-898716-31-3

90000

9 780898 716313