Princeton Series in APPLIED MATHEMATICS

Positive Definite Matrices

Rajendra Bhatia

Positive Definite Matrices

Positive Definite Matrices

Rajendra Bhatia

PRINCETON UNIVERSITY PRESS
PRINCETON AND OXFORD

Copyright ©2007 by Princeton University Press

Requests for permission to reproduce material from this work
should be sent to Permissions, Princeton University Press

Published by Princeton University Press,
41 William Street, Princeton, New Jersey 08540
In the United Kingdom: Princeton University Press,
3 Market Place, Woodstock, Oxfordshire OX20 1SY

All Rights Reserved

Library of Congress Cataloging-in-Publication Data

Bhatia, Rajendra, 1952-
Positive definite matrices / Rajendra Bhatia
p. cm. Includes bibliographical references and index.
ISBN-13: 978-0-691-12918-1 (cloth : alk. paper)
ISBN-10: 0-691-12918-5 (cloth : alk. paper)
1. Matrices. I. Title.
QA188.B488 2007
512.9'434–dc22 2006050375

British Library Cataloging-in-Publication Data is available

The publisher would like to acknowledge the author of this
volume for providing the camera-ready copy from which this book was printed.

This book has been composed in Times Roman using LaTeX

Printed on acid-free paper. ∞

pup.princeton.edu

Printed in the United States of America

1 3 5 7 9 10 8 6 4 2

Contents

Preface

The theory of positive definite matrices, positive definite functions, and positive linear maps is rich in content. It offers many beautiful theorems that are simple and yet striking in their formulation, uncomplicated and yet ingenious in their proof, diverse as well as powerful in their application. The aim of this book is to present some of these results with a minimum of prerequisite preparation.

The seed of this book lies in a cycle of lectures I was invited to give at the Centro de Estruturas Lineares e Combinatórias (CELC) of the University of Lisbon in the summer of 2001. My audience was made up of seasoned mathematicians with a distinguished record of research in linear and multilinear algebra, combinatorics, group theory, and number theory. The aim of the lectures was to draw their attention to some results and methods used by analysts. A preliminary draft of the first four chapters was circulated as lecture notes at that time. Chapter 5 was added when I gave another set of lectures at the CELC in 2003.

Because of this genesis, the book is oriented towards those interested in linear algebra and matrix analysis. In some ways it supplements my earlier book *Matrix Analysis* (Springer, Graduate Texts in Mathematics, Volume 169). However, it can be read independently of that book. The usual graduate-level preparation in analysis, functional analysis, and linear algebra provides adequate background needed for reading this book.

Chapter 1 contains some basic ideas used throughout the book. Among other things it introduces the reader to some arguments involving 2×2 block matrices. These have been used to striking, almost magical, effect by T. Ando, M.-D. Choi, and other masters of the subject and the reader will see some of that in later parts of the book.

Chapters 2 and 3 are devoted to the study of positive and completely positive maps with special emphasis on their use in proving matrix inequalities. Most of this material is very well known to those who study C^*-algebras, and it ought to be better known to workers in linear algebra. In the book, as in my Lisbon lectures, I have avoided the technical difficulties of the theory of operator algebras by staying in finite-dimensional spaces. Thus some of the major theorems of the subject are presented in their toy versions. This is good enough for the purposes of matrix analysis and also of the currently popular area of quantum information theory. Quantum communication channels, at present, are thought of as completely positive trace-preserving linear maps on matrix algebras and many problems of the subject are phrased in terms of block matrices.

In Chapter 4 we discuss means of two positive definite matrices with special emphasis on the geometric mean. Among spectacular applications of these ideas we include proofs of some theorems on matrix convex functions, and of two of the most famous theorems on quantum mechanical entropy.

Chapter 5 gives a quick introduction to positive definite functions on the real line. Many examples of such functions are constructed using elementary arguments and then used to derive matrix inequalities. Again, a special emphasis has been placed on various means of matrices. Many of the results presented are drawn from recent research work.

Chapter 6 is, perhaps, somewhat unusual. It presents some standard and important theorems of Riemannian geometry as seen from the perspective of matrix analysis. Positive definite matrices constitute a Riemannian manifold of nonpositive curvature, a much-studied object in differential geometry. After explaining the basic ideas in a language more familiar to analysts we show how these are used to define geometric means of more than two matrices. Such a definition has been elusive for long and only recently some progress has been made. It leads to some intriguing questions for both the analyst and the geometer.

This is neither an encyclopedia nor a compendium of all that is known about positive definite matrices. It is possible to use this book for a one semester topics course at the graduate level. Several exercises of varying difficulty are included and some research problems are mentioned. Each chapter ends with a section called "Notes and References". Again, these are written to inform certain groups of readers, and are not intended to be scholarly commentaries.

The phrase positive matrix has been used all through the book to mean a positive semidefinite, or a positive definite, matrix. No confusion should be caused by this. Occasionally I refer to my book *Matrix Analysis.* Most often this is done to recall some standard result. Sometimes I do it to make a tangential point that may be ignored without losing anything of the subsequent discussion. In each case a reference like "MA, page xx" or "See Section x.y.z of MA" points to the relevant page or section of *Matrix Analysis.*

Over the past 25 years I have learnt much from several colleagues and friends. I was a research associate of W. B. Arveson at Berkeley in 1979–80, of C. Davis and M.-D. Choi at Toronto in 1983, and of T. Ando at Sapporo in 1985. This experience has greatly influenced my work and my thinking and I hope some of it is reflected in this book. I have had a much longer, and a more constant, association with K. R. Parthasarathy. Chapter 5 of the book is based on work I did with him and the understanding I obtained during the process. Likewise Chapter 6 draws on the efforts J.A.R. Holbrook and I together made to penetrate the mysteries of territory not familiar to us.

D. Drissi, L. Elsner, R. Horn, F. Kittaneh, K. B. Sinha, and X. Zhan have been among my frequent collaborators and correspondents and have generously shared their ideas and insights with me. F. Hiai and H. Kosaki have often sent me their papers before publication, commented on my work, and clarified many issues about which I have written here. In particular, Chapter 5 contains many of their ideas.

My visits to Lisbon were initiated and organized by J. A. Dias da Silva and F. C. Silva. I was given a well-appointed office, a good library, and a comfortable apartment—all within 20 meters of each other, a faithful and devoted audience for my lectures, and a cheerful and competent secretary to type my notes. In these circumstances it would have been extraordinarily slothful not to produce a book.

The hard work and good cheer of Fernanda Proença at the CELC were continued by Anil Shukla at the Indian Statistical Institute, Delhi. Between the two of them several drafts of the book have been processed over a period of five years.

Short and long lists of minor and major mistakes in the evolving manuscript were provided by helpful colleagues: they include J. S. Aujla, J. C. Bourin, A. Dey, B. P. Duggal, T. Furuta, F. Hiai, J.A.R. Holbrook, M. Moakher, and A. I. Singh. But even their hawk eyes might have missed some bugs. I can only hope these are both few and benignant.

I am somewhat perplexed by authors who use this space to suggest that their writing activities cause acute distress to their families and to thank them for bearing it all in the cause of humanity. My wife Irpinder and son Gautam do deserve thanks, but my writing does not seem to cause them any special pain.

It is a pleasure to record my thanks to all the individuals and institutions named above.

Chapter One

Positive Matrices

We begin with a quick review of some of the basic properties of positive matrices. This will serve as a warmup and orient the reader to the line of thinking followed through the book.

1.1 CHARACTERIZATIONS

Let $\mathcal{H}$ be the n-dimensional Hilbert space $\mathbb{C}^n$. The inner product between two vectors x and y is written as $\langle x, y\rangle$ or as x^*y. We adopt the convention that the inner product is conjugate linear in the first variable and linear in the second. We denote by $\mathcal{L}(\mathcal{H})$ the space of all linear operators on $\mathcal{H}$, and by $\mathbb{M}_n(\mathbb{C})$ or simply $\mathbb{M}_n$ the space of $n \times n$ matrices with complex entries. Every element A of $\mathcal{L}(\mathcal{H})$ can be identified with its matrix with respect to the standard basis $\{e_j\}$ of $\mathbb{C}^n$. We use the symbol A for this matrix as well. We say A is *positive semidefinite* if

$$\langle x, Ax\rangle \quad \geq \quad 0 \quad \text{for all} \quad x \in \mathcal{H}, \tag{1.1}$$

and *positive definite* if, in addition,

$$\langle x, Ax\rangle \quad > \quad 0 \quad \text{for all} \quad x \neq 0. \tag{1.2}$$

A positive semidefinite matrix is positive definite if and only if it is invertible.

For the sake of brevity, we use the term *positive* matrix for a positive semidefinite, or a positive definite, matrix. Sometimes, if we want to emphasize that the matrix is positive definite, we say that it is *strictly positive.* We use the notation $A \geq O$ to mean that A is positive, and $A > O$ to mean it is strictly positive.

There are several conditions that characterize positive matrices. Some of them are listed below.

(i) A is positive if and only if it is Hermitian ($A = A^*$) and all its eigenvalues are nonnegative. A is strictly positive if and only if all its eigenvalues are positive.

(ii) A is positive if and only if it is Hermitian and all its principal minors are nonnegative. A is strictly positive if and only if all its principal minors are positive.

(iii) A is positive if and only if $A = B^*B$ for some matrix B. A is strictly positive if and only if B is nonsingular.

(iv) A is positive if and only if $A = T^*T$ for some upper triangular matrix T. Further, T can be chosen to have nonnegative diagonal entries. If A is strictly positive, then T is unique. This is called the *Cholesky decomposition* of A. A is strictly positive if and only if T is nonsingular.

(v) A is positive if and only if $A = B^2$ for some positive matrix B. Such a B is unique. We write $B = A^{1/2}$ and call it the (positive) square root of A. A is strictly positive if and only if B is strictly positive.

(vi) A is positive if and only if there exist $x_1, \ldots, x_n$ in $\mathcal{H}$ such that

$$a_{ij} = \langle x_i, x_j \rangle. \tag{1.3}$$

A is strictly positive if and only if the vectors x_j, $1 \le j \le n$, are linearly independent.

A proof of the sixth characterization is outlined below. This will serve the purpose of setting up some notations and of introducing an idea that will be often used in the book.

We think of elements of $\mathbb{C}^n$ as column vectors. If $x_1, \ldots, x_m$ are such vectors we write $[x_1, \ldots, x_m]$ for the $n \times m$ matrix whose columns are $x_1, \ldots, x_m$. The adjoint of this matrix is written as

$$\begin{bmatrix} x_1^* \\ \vdots \\ x_m^* \end{bmatrix}.$$

This is an $m \times n$ matrix whose rows are the (row) vectors $x_1^*, \ldots, x_m^*$. We use the symbol $[[a_{ij}]]$ for a matrix with i, j entry a_{ij}.

Now if $x_1, \ldots, x_n$ are elements of $\mathbb{C}^n$, then

$$[[x_i^* x_j]] = \begin{bmatrix} x_1^* \\ \vdots \\ x_n^* \end{bmatrix} [x_1, \dots, x_n].$$

So, this matrix is positive (being of the form B^*B). This shows that the condition (1.3) is sufficient for A to be positive. Conversely, if A is positive, we can write

$$a_{ij} = \langle e_i, Ae_j \rangle = \langle A^{1/2} e_i, A^{1/2} e_j \rangle.$$

If we choose $x_j = A^{1/2} e_j$, we get (1.3).

1.1.1 Exercise

Let $x_1, \dots, x_m$ be any m vectors in *any* Hilbert space. Then the $m \times m$ matrix

$$G(x_1, \dots, x_m) = [[x_i^* x_j]] \tag{1.4}$$

is positive. It is strictly positive if and only if $x_1, \dots, x_m$ are linearly independent.

The matrix (1.4) is called the *Gram matrix* associated with the vectors $x_1, \dots, x_m$.

1.1.2 Exercise

Let $\lambda_1, \dots, \lambda_m$ be positive numbers. The $m \times m$ matrix A with entries

$$a_{ij} = \frac{1}{\lambda_i + \lambda_j} \tag{1.5}$$

is called the Cauchy matrix (associated with the numbers λ_j). Note that

$$a_{ij} = \int_0^\infty e^{-(\lambda_i + \lambda_j)t} dt. \tag{1.6}$$

Let $f_i(t) = e^{-\lambda_i t}$, $1 \le i \le m$. Then $f_i \in L_2([0, \infty))$ and $a_{ij} = \langle f_i, f_j \rangle$. This shows that A is positive.

More generally, let $\lambda_1, \ldots, \lambda_m$ be complex numbers whose real parts are positive. Show that the matrix A with entries

$$a_{ij} = \frac{1}{\overline{\lambda}_i + \lambda_j}$$

is positive.

1.1.3 Exercise

Let μ be a finite positive measure on the interval $[-\pi, \pi]$. The numbers

$$a_m = \int_{-\pi}^{\pi} e^{-im\theta} d\mu(\theta), \qquad m \in \mathbb{Z}, \tag{1.7}$$

are called the *Fourier-Stieltjes coefficients* of μ. For any $n = 1, 2, \ldots,$ let A be the $n \times n$ matrix with entries

$$\alpha_{ij} = a_{i-j}, \qquad 0 \leq i, j \leq n-1. \tag{1.8}$$

Then A is positive.

Note that the matrix A has the form

$$A = \begin{bmatrix} a_0 & \overline{a}_1 & \ldots & \overline{a}_{n-1} \\ a_1 & a_0 & \overline{a}_1 \ddots & \\ & & & \\ \vdots & \ddots & \ddots & \overline{a}_1 \\ a_{n-1} & \ldots & a_1 & a_0 \end{bmatrix}. \tag{1.9}$$

One special feature of this matrix is that its entries are constant along the diagonals parallel to the main diagonal. Such a matrix is called a *Toeplitz matrix.* In addition, A is Hermitian.

A doubly infinite sequence $\{a_m : m \in \mathbb{Z}\}$ of complex numbers is said to be a *positive definite sequence* if for each $n = 1, 2, \ldots,$ the $n \times n$ matrix (1.8) constructed from this sequence is positive. We have seen that the Fourier-Stieltjes coefficients of a finite positive measure on $[-\pi, \pi]$ form a positive definite sequence. A basic theorem of harmonic analysis called the Herglotz theorem says that, conversely, every positive definite sequence is the sequence of Fourier-Stieltjes coefficients of a finite positive measure μ. This theorem is proved in Chapter 5.

1.2 SOME BASIC THEOREMS

Let A be a positive operator on $\mathcal{H}$. If X is a linear operator from a Hilbert space $\mathcal{K}$ into $\mathcal{H}$, then the operator X^*AX on $\mathcal{K}$ is also positive. If X is an invertible opertator, and X^*AX is positive, then A is positive.

Let A, B be operators on $\mathcal{H}$. We say that A is *congruent* to B, and write $A \sim B$, if there exists an invertible operator X on $\mathcal{H}$ such that $B = X^*AX$. Congruence is an equivalence relation on $\mathcal{L}(\mathcal{H})$. If X is unitary, we say A is *unitarily equivalent* to B, and write $A \simeq B$.

If A is Hermitian, the *inertia* of A is the triple of nonnegative integers

$$\mathrm{In}(A) = (\pi(A), \zeta(A), \nu(A)), \tag{1.10}$$

where $\pi(A)$, $\zeta(A)$, $\nu(A)$ are the numbers of positive, zero, and negative eigenvalues of A (counted with multiplicity).

Sylvester's law of inertia says that $\mathrm{In}(A)$ is a complete invariant for congruence on the set of Hermitian matrices; i.e., two Hermitian matrices are congruent if and only if they have the same inertia. This can be proved in different ways. Two proofs are outlined below.

1.2.1 Exercise

Let $\lambda_1^{\downarrow}(A) \geq \cdots \geq \lambda_n^{\downarrow}(A)$ be the eigenvalues of the Hermitian matrix A arranged in decreasing order. By the minimax principle (MA, Corollary III.1.2)

$$\lambda_k^{\downarrow}(A) = \max_{\dim \mathcal{M}=k} \quad \min_{\substack{x \in \mathcal{M} \\ \|x\|=1}} \quad \langle x, Ax \rangle,$$

where $\mathcal{M}$ stands for a subspace of $\mathcal{H}$ and $\dim \mathcal{M}$ for its dimension. If X is an invertible operator, then $\dim X(\mathcal{M}) = \dim \mathcal{M}$. Use this to prove that any two congruent Hermitian matrices have the same inertia.

1.2.2 Exercise

Let A be a nonsingular Hermitian matrix and let $B = X^*AX$, where X is any nonsingular matrix. Let X have the polar decomposition $X = UP$, where U is unitary and P is strictly positive. Let

$$\begin{aligned} P(t) &= (1-t)I + tP, & 0 \le t \le 1\,, \\ X(t) &= UP(t), & 0 \le t \le 1\,. \end{aligned}$$

Then $P(t)$ is strictly positive, and $X(t)$ nonsingular. We have $X(0) = U$, and $X(1) = X$. Thus $X(t)^*AX(t)$ is a continuous curve in the space of nonsingular matrices joining U^*AU and X^*AX. The eigenvalues of $X(t)^*AX(t)$ are continuous curves joining the eigenvalues of U^*AU (these are the same as the eigenvalues of A) and the eigenvalues of $X^*AX = B$. [MA, Corollary VI.1.6]. These curves never touch the point zero. Hence

$$\pi(A) = \pi(B); \quad \zeta(A) = \zeta(B) = 0; \quad \nu(A) = \nu(B);$$

i.e., A and B have the same inertia.
Modify this argument to cover the case when A is singular. (Then $\zeta(A) = \zeta(B)$. Consider $A \pm \varepsilon I$.)

1.2.3 Exercise

Show that a Hermitian matrix A is congruent to the diagonal matrix diag $(1, \ldots, 1, 0, \ldots, 0, -1, \ldots, -1)$, in which the entries $1, 0, -1$ occur $\pi(A)$, $\zeta(A)$, and $\nu(A)$ times on the diagonal. Thus two Hermitian matrices with the same inertia are congruent.

Two Hermitian matrices are unitarily equivalent if and only if they have the same eigenvalues (counted with multiplicity).

Let $\mathcal{K}$ be a subspace of $\mathcal{H}$ and let P be the orthogonal projection onto $\mathcal{K}$. If we choose an orthonormal basis in which $\mathcal{K}$ is spanned by the first k vectors, then we can write an operator A on $\mathcal{H}$ as a block matrix

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

and

$$PAP = \begin{bmatrix} A_{11} & O \\ O & O \end{bmatrix}.$$

If V is the injection of $\mathcal{K}$ into $\mathcal{H}$, then $V^*AV = A_{11}$. We say that A_{11} is the *compression* of A to $\mathcal{K}$.

If A is positive, then all its compressions are positive. Thus all principal submatrices of a positive matrix are positive. Conversely, if all the principal subdeterminants of A are nonnegative, then the coefficients in the characteristic polynomial of A alternate in sign. Hence, by the Descartes rule of signs A has no negative root.

The following exercise says that if all the *leading* subdeterminants of a Hermitian matrix A are positive, then A is strictly positive. Positivity of other principal minors follows as a consequence.

Let A be Hermitian and let B be its compression to an $(n-k)$-dimensional subspace. Then Cauchy's interlacing theorem [MA, Corollary III.1.5] says that

$$\lambda_j^{\downarrow}(A) \geq \lambda_j^{\downarrow}(B) \geq \lambda_{j+k}^{\downarrow}(A). \tag{1.11}$$

1.2.4 Exercise

(i) If A is strictly positive, then all its compressions are strictly positive.

(ii) For $1 \leq j \leq n$ let $A_{[j]}$ denote the $j \times j$ block in the top left corner of the matrix of A. Call this the *leading* $j \times j$ submatrix of A, and its determinant the leading $j \times j$ subdeterminant. Show that A is strictly positive if and only if all its leading subdeterminants are positive. [Hint: use induction and Cauchy's interlacing theorem.]

The example $A = \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$ shows that nonnegativity of the two leading subdeterminants is not adequate to ensure positivity of A.

We denote by $A \otimes B$ the tensor product of two operators A and B (acting possibly on different Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$). If A, B are positive, then so is $A \otimes B$.

If A, B are $n \times n$ matrices we write $A \circ B$ for their entrywise product; i.e., for the matrix whose i, j entry is $a_{ij}b_{ij}$. We will call this the *Schur product* of A and B. It is also called the *Hadamard product.* If A and B are positive, then so is $A \circ B$. One way of seeing this is by observing that $A \circ B$ is a principal submatrix of $A \otimes B$.

1.2.5 Exercise

Let A, B be positive matrices of rank one. Then there exist vectors x, y such that $A = xx^*$, $B = yy^*$. Show that $A \circ B = zz^*$, where z is the vector $x \circ y$ obtained by taking entrywise product of the coordinates of x and y. Thus $A \circ B$ is positive. Use this to show that the Schur product of any two positive matrices is positive.

If both A, B are Hermitian, or positive, then so is $A + B$. Their product AB is, however, Hermitian if and only if A and B commute. This condition is far too restrictive. The *symmetrized product* of A, B is the matrix

$$S = AB + BA. \tag{1.12}$$

If A, B are Hermitian, then S is Hermitian. However, if A, B are positive, then S need not be positive. For example, the matrices

$$A = \begin{bmatrix} 1 & 0 \\ 0 & \varepsilon \end{bmatrix}, \quad B = \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix}$$

are positive if $\varepsilon > 0$ and $0 < \alpha < 1$, but S is not positive when ε is close to zero and α is close to 1. In view of this it is, perhaps, surprising that if S is positive and A strictly positive, then B is positive. Three different proofs of this are outlined below.

1.2.6 Proposition

Let A, B be Hermitian and suppose A is strictly positive. If the symmetrized product $S = AB + BA$ is positive (strictly positive), then B is positive (strictly positive).

Proof. Choose an orthonormal basis in which B is diagonal; $B = \text{diag}(\beta_1, \ldots, \beta_n)$. Then $s_{ii} = 2\beta_i a_{ii}$. Now observe that the diagonal entries of a (strictly) positive matrix are (strictly) positive. ■

1.2.7 Exercise

Choose an orthonormal basis in which A is diagonal with entries $\alpha_1, \alpha_2, \ldots, \alpha_n$, on its diagonal. Then note that S is the Schur product of B with the matrix $[[\alpha_i + \alpha_j]]$. Hence B is the Schur product of S with the Cauchy matrix $[[1/(\alpha_i + \alpha_j)]]$. Since this matrix is positive, it follows that B is positive if S is.

1.2.8 Exercise

If $S > O$, then for every nonzero vector x

$$0 < \langle x, Sx \rangle = 2 \operatorname{Re} \langle x, ABx \rangle.$$

Suppose $Bx = \beta x$ with $\beta \leq 0$. Show that $\langle x, ABx \rangle \leq 0$. Conclude that $B > O$.

An amusing corollary of Proposition 1.2.6 is a simple proof of the operator monotonicity of the map $A \longmapsto A^{1/2}$ on positive matrices.

If A, B are Hermitian, we say that $A \geq B$ if $A - B \geq O$; and $A > B$ if $A - B > O$.

1.2.9 Proposition

If A, B are positive and $A > B$, then $A^{1/2} > B^{1/2}$.

Proof. We have the identity

$$X^2 - Y^2 = \frac{(X+Y)(X-Y) + (X-Y)(X+Y)}{2}. \tag{1.13}$$

If X, Y are strictly positive then $X + Y$ is strictly positive. So, if $X^2 - Y^2$ is positive, then $X - Y$ is positive by Proposition 1.2.6. ■

Recall that if $A \geq B$, then we need not always have $A^2 \geq B^2$; e.g., consider the matrices

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Proposition 1.2.6 is related to the study of the *Lyapunov equation*, of great importance in differential equations and control theory. This is the equation (in matrices)

$$A^* X + XA = W. \tag{1.14}$$

It is assumed that the spectrum of A is contained in the open right half-plane. The matrix A is then called *positively stable*. It is well known that in this case the equation (1.14) has a unique solution. Further, if W is positive, then the solution X is also positive.

1.2.10 Exercise

Suppose A is diagonal with diagonal entries $\alpha_1, \ldots, \alpha_n$. Then the solution of (1.14) is

$$X = \left[\left[\frac{1}{\overline{\alpha}_\imath + \alpha_j}\right]\right] \circ W.$$

Use Exercise 1.1.2 to see that if W is positive, then so is X. Now suppose $A = TDT^{-1}$, where D is diagonal. Show that again the solution X is positive if W is positive. Since diagonalisable matrices are dense in the space of all matrices, the same conclusion can be obtained for general positively stable A.

The solution X to the equation (1.14) can be represented as the integral

$$X = \int_0^\infty e^{-tA^*} W e^{-tA} dt. \tag{1.15}$$

The condition that A is positively stable ensures that this integral is convergent. It is easy to see that X defined by (1.15) satisfies the equation (1.14). From this it is clear that if W is positive, then so is X.

Now suppose A is any matrix and suppose there exist positive matrices X and W such that the equality (1.14) holds. Then if $Au = \alpha u$, we have

$$\begin{aligned}\langle u, Wu\rangle &= \langle u, (A^*X + XA)u\rangle = \langle XAu, u\rangle + \langle u, XAu\rangle \\ &= 2\,\mathrm{Re}\,\alpha\langle Xu, u\rangle.\end{aligned}$$

This shows that A is positively stable.

1.2.11 Exercise

The matrix equation

$$X - F^*XF = W \tag{1.16}$$

is called the Stein equation or the discrete time Lyapunov equation. It is assumed that the spectrum of F is contained in the open unit disk. Show that in this case the equation has a unique solution given by

$$X = \sum_{m=0}^{\infty} F^{*m} W F^m. \tag{1.17}$$

From this it is clear that if W is positive, then so is X. Another proof of this fact goes as follows. To each point β in the open unit disk there corresponds a unique point α in the open right half plane given by $\beta = \frac{\alpha-1}{\alpha+1}$. Suppose F is diagonal with diagonal entries $\beta_1, \ldots, \beta_n$. Then the solution of (1.16) can be written as

$$X = \left[\left[\frac{1}{1-\overline{\beta}_i\beta_j}\right]\right] \circ W.$$

Use the correspondence between β and α to show that

$$\left[\left[\frac{1}{1-\overline{\beta}_i\beta_j}\right]\right] = \left[\left[\frac{(\overline{\alpha}_i+1)(\alpha_j+1)}{2(\overline{\alpha}_i+\alpha_j)}\right]\right] \sim \left[\left[\frac{1}{\overline{\alpha}_i+\alpha_j}\right]\right].$$

Now use Exercise 1.2.10.

If F is any matrix such that the equality (1.16) is satisfied by some positive matrices X and W, then the spectrum of F is contained in the unit disk.

1.2.12 Exercise

Let A, B be strictly positive matrices such that $A \geq B$. Show that $A^{-1} \leq B^{-1}$. [Hint: If $A \geq B$, then $I \geq A^{-1/2}BA^{-1/2}$.]

1.2.13 Exercise

The quadratic equation

$$XAX = B$$

is called a *Riccati equation.* If B is positive and A strictly positive, then this equation has a positive solution. Conjugate the two sides of the equation by $A^{1/2}$, take square roots, and then conjugate again by $A^{-1/2}$ to see that

$$X = A^{-1/2}(A^{1/2}BA^{1/2})^{1/2}A^{-1/2}$$

is a solution. Show that this is the only positive solution.

1.3 BLOCK MATRICES

Now we come to a major theme of this book. We will see that 2×2 block matrices $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ can play a remarkable—almost magical—role in the study of positive matrices.

In this block matrix the entries A, B, C, D are $n \times n$ matrices. So, the big matrix is an element of $\mathbb{M}_{2n}$, or, of $\mathcal{L}(\mathcal{H} \oplus \mathcal{H})$. As we proceed we will see that several properties of A can be obtained from those of a block matrix in which A is one of the entries. Of special importance is the connection this establishes between positivity (an algebraic property) and contractivity (a metric property).

Let us fix some notations. We will write $A = UP$ for the *polar decomposition* of A. The factor U is unitary and P is positive; we have $P = (A^*A)^{1/2}$. This is called the *positive part* or the *absolute value* of A and is written as $|A|$. We have $A^* = PU^*$, and

$$|A^*| = (AA^*)^{1/2} = (UP^2U^*)^{1/2} = UPU^*.$$

A is said to be normal if $AA^* = A^*A$. This condition is equivalent to $UP = PU$; and to the condition $|A| = |A^*|$.

We write $A = USV$ for the *singular value decomposition* (SVD) of A. Here U and V are unitary and S is diagonal with nonnegative diagonal entries $s_1(A) \geq \cdots \geq s_n(A)$. These are the singular values of A (the eigenvalues of $|A|$).

The symbol $\|A\|$ will always denote the norm of A as a linear operator on the Hilbert space $\mathcal{H}$; i.e.,

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

It is easy to see that $\|A\| = s_1(A)$.

Among the important properties of this norm are the following:

$$\begin{aligned} \|AB\| &\leq \|A\|\|B\|, \\ \|A\| &= \|A^*\|, \\ \|A\| &= \|UAV\| \text{ for all unitary } U, V. \end{aligned} \tag{1.18}$$

This last property is called *unitary invariance*. Finally

$$\|A^*A\| = \|A\|^2. \tag{1.19}$$

There are several other norms on $\mathbb{M}_n$ that share the three properties (1.18). It is the condition (1.19) that makes the operator norm $\|\cdot\|$ very special.

We say A is *contractive*, or A is a *contraction*, if $\|A\| \le 1$.

1.3.1 Proposition

The operator A is contractive if and only if the operator $\begin{bmatrix} I & A \\ A^* & I \end{bmatrix}$ is positive.

Proof. What does the proposition say when $\mathcal{H}$ is one-dimensional? It just says that if a is a complex number, then $|a| \le 1$ if and only if the 2×2 matrix $\begin{bmatrix} 1 & a \\ \bar{a} & 1 \end{bmatrix}$ is positive. The passage from one to many dimensions is made via the SVD. Let $A = USV$. Then

$$\begin{aligned} \begin{bmatrix} I & A \\ A^* & I \end{bmatrix} &= \begin{bmatrix} I & USV \\ V^*SU^* & I \end{bmatrix} \\ &= \begin{bmatrix} U & O \\ O & V^* \end{bmatrix} \begin{bmatrix} I & S \\ S & I \end{bmatrix} \begin{bmatrix} U^* & O \\ O & V \end{bmatrix}. \end{aligned}$$

This matrix is unitarily equivalent to $\begin{bmatrix} I & S \\ S & I \end{bmatrix}$, which in turn is unitarily equivalent to the direct sum

$$\begin{bmatrix} 1 & s_1 \\ s_1 & 1 \end{bmatrix} \oplus \begin{bmatrix} 1 & s_2 \\ s_2 & 1 \end{bmatrix} \oplus \cdots \oplus \begin{bmatrix} 1 & s_n \\ s_n & 1 \end{bmatrix},$$

where $s_1, \ldots, s_n$ are the singular values of A. These 2×2 matrices are all positive if and only if $s_1 \le 1$ (i.e.,$\|A\| \le 1$). ■

1.3.2 Proposition

Let A, B be positive. Then the matrix $\begin{bmatrix} A & X \\ X^* & B \end{bmatrix}$ is positive if and only if $X = A^{1/2}KB^{1/2}$ for some contraction K.

Proof. Assume first that A, B are strictly positive. This allows us to use the congruence

$$\begin{bmatrix} A & X \\ X^* & B \end{bmatrix} \sim \begin{bmatrix} A^{-1/2} & O \\ O & B^{-1/2} \end{bmatrix} \begin{bmatrix} A & X \\ X^* & B \end{bmatrix} \begin{bmatrix} A^{-1/2} & O \\ O & B^{-1/2} \end{bmatrix}$$

$$= \begin{bmatrix} I & A^{-1/2}XB^{-1/2} \\ B^{-1/2}X^*A^{-1/2} & I \end{bmatrix}.$$

Let $K = A^{-1/2}XB^{-1/2}$. Then by Proposition 1.3.1 this block matrix is positive if and only if K is a contraction. This proves the proposition when A, B are strictly positive. The general case follows by a continuity argument. ■

It follows from Proposition 1.3.2 that if $\begin{bmatrix} A & X \\ X^* & B \end{bmatrix}$ is positive, then the range of X is a subspace of the range of A, and the range of X^* is a subspace of the range of B. The rank of X cannot exceed either the rank of A or the rank of B.

1.3.3 Theorem

Let A, B be strictly positive matrices. Then the block matrix $\begin{bmatrix} A & X \\ X^* & B \end{bmatrix}$ is positive if and only if $A \geq XB^{-1}X^*$.

Proof. We have the congruence

$$\begin{bmatrix} A & X \\ X^* & B \end{bmatrix} \sim \begin{bmatrix} I & -XB^{-1} \\ O & I \end{bmatrix} \begin{bmatrix} A & X \\ X^* & B \end{bmatrix} \begin{bmatrix} I & O \\ -B^{-1}X^* & I \end{bmatrix}$$
$$= \begin{bmatrix} A - XB^{-1}X^* & O \\ O & B \end{bmatrix}.$$

Clearly, this last matrix is positive if and only if $A \geq XB^{-1}X^*$. ■

Second proof. We have $A \geq XB^{-1}X^*$ if and only if

$$\begin{aligned} I &\geq A^{-1/2}(XB^{-1}X^*)A^{-1/2} \\ &= A^{-1/2}XB^{-1/2} \cdot B^{-1/2}X^*A^{-1/2} \\ &= (A^{-1/2}XB^{-1/2})(A^{-1/2}XB^{-1/2})^*. \end{aligned}$$

This is equivalent to saying $\|A^{-1/2}XB^{-1/2}\| \leq 1$, or $X = A^{1/2}KB^{1/2}$ where $\|K\| \leq 1$. Now use Proposition 1.3.2. ■

1.3.4 Exercise

Show that the condition $A \geq XB^{-1}X^*$ in the theorem cannot be replaced by $A \geq X^*B^{-1}X$ (except when $\mathcal{H}$ is one dimensional!).

1.3.5 Exercise

Let A, B be positive but not strictly positive. Show that Theorem 1.3.3 is still valid if B^{-1} is interpreted to be the Moore-Penrose inverse of B.

1.3.6 Lemma

The matrix A is positive if and only if $\begin{bmatrix} A & A \\ A & A \end{bmatrix}$ is positive.

Proof. We can write

$$\begin{bmatrix} A & A \\ A & A \end{bmatrix} = \begin{bmatrix} A^{1/2} & O \\ A^{1/2} & O \end{bmatrix} \begin{bmatrix} A^{1/2} & A^{1/2} \\ O & O \end{bmatrix}.$$

1.3.7 Corollary

Let A be any matrix. Then the matrix $\begin{bmatrix} |A| & A^* \\ A & |A^*| \end{bmatrix}$ is positive.

Proof. Use the polar decomposition $A = UP$ to write

$$\begin{bmatrix} |A| & A^* \\ A & |A^*| \end{bmatrix} = \begin{bmatrix} P & PU^* \\ UP & UPU^* \end{bmatrix}$$
$$= \begin{bmatrix} I & O \\ O & U \end{bmatrix} \begin{bmatrix} P & P \\ P & P \end{bmatrix} \begin{bmatrix} I & O \\ O & U^* \end{bmatrix},$$

and then use the lemma. ■

1.3.8 Corollary

If A is normal, then $\begin{bmatrix} |A| & A^* \\ A & |A| \end{bmatrix}$ is positive.

1.3.9 Exercise

Show that (when $n \geq 2$) this is not always true for nonnormal matrices.

1.3.10 Exercise

If A is strictly positive, show that $\begin{bmatrix} A & I \\ I & A^{-1} \end{bmatrix}$ is positive (but not strictly positive.)

Theorem 1.3.3, the other propositions in this section, and the ideas used in their proofs will occur repeatedly in this book. Some of their power is demonstrated in the next sections.

1.4 NORM OF THE SCHUR PRODUCT

Given A in $\mathbb{M}_n$, let S_A be the linear map on $\mathbb{M}_n$ defined as

$$S_A(X) = A \circ X, \quad X \in \mathbb{M}_n, \tag{1.20}$$

where $A \circ X$ is the Schur product of A and X. The norm of this linear operator is, by definition,

$$\|S_A\| = \sup_{\|X\|=1} \|S_A(X)\| = \sup_{\|X\|\leq 1} \|S_A(X)\|. \tag{1.21}$$

Since

$$\|A \circ B\| \leq \|A \otimes B\| = \|A\| \, \|B\|, \tag{1.22}$$

we have

$$\|S_A\| \leq \|A\|. \tag{1.23}$$

Finding the exact value of $\|S_A\|$ is a difficult problem in general. Some special cases are easier.

1.4.1 Theorem (Schur)

If A is positive, then

$$\|S_A\| = \max \, a_{ii}. \tag{1.24}$$

Proof. Let $\|X\| \leq 1$. Then by Proposition 1.3.1 $\begin{bmatrix} I & X \\ X^* & I \end{bmatrix} \geq O$. By Lemma 1.3.6 $\begin{bmatrix} A & A \\ A & A \end{bmatrix} \geq O$. Hence the Schur product of these two block matrices is positive; i.e,

$$\begin{bmatrix} A \circ I & A \circ X \\ (A \circ X)^* & A \circ I \end{bmatrix} \geq O.$$

So, by Proposition 1.3.2, $\|A \circ X\| \leq \|A \circ I\| = \max\ a_{ii}$. Thus $\|S_A\| = \max\ a_{ii}$. ■

1.4.2 Exercise

If U is unitary, then $\|S_U\| = 1$. [Hint: $U \circ \overline{U}$ is doubly stochastic, and hence, has norm 1.]

For each matrix X, let

$$\|X\|_c \ = \text{maximum of the Euclidean norms of columns of } X. \quad (1.25)$$

This is a norm on $\mathbb{M}_n$, and

$$\|X\|_c \leq \|X\|. \quad (1.26)$$

1.4.3 Theorem

Let A be any matrix. Then

$$\|S_A\| \leq \ \inf \{\|X\|_c \ \|Y\|_c \ : \ A = X^*Y\}. \quad (1.27)$$

Proof. Let $A = X^*Y$. Then

$$\begin{bmatrix} X^*X & A \\ A^* & Y^*Y \end{bmatrix} = \begin{bmatrix} X^* & O \\ Y^* & O \end{bmatrix} \begin{bmatrix} X & Y \\ O & O \end{bmatrix} \geq O. \quad (1.28)$$

Now if Z is any matrix with $\|Z\| \leq 1$, then $\left[\begin{smallmatrix} I & Z \\ Z^* & I \end{smallmatrix}\right] \geq O$. So, the Schur product of this with the positive matrix in (1.28) is positive; i.e.,

$$\begin{bmatrix} (X^*X) \circ I & A \circ Z \\ (A \circ Z)^* & (Y^*Y) \circ I \end{bmatrix} \geq O.$$

Hence, by Proposition 1.3.2

$$\|A \circ Z\| \leq \|(X^*X) \circ I\|^{1/2} \ \ \|(Y^*Y) \circ I\|^{1/2} = \|X\|_c \|Y\|_c.$$

Thus $\|S_A\| \leq \|X\|_c \|Y\|_c$. ■

In particular, we have

$$\|S_A\| \le \|A\|_c, \tag{1.29}$$

which is an improvement on (1.23).

In Chapter 3, we will prove a theorem of Haagerup (Theorem 3.4.3) that says the two sides of (1.27) are equal.

1.5 MONOTONICITY AND CONVEXITY

Let $\mathcal{L}_{s.a.}(\mathcal{H})$ be the space of self-adjoint (Hermitian) operators on $\mathcal{H}$. This is a real vector space. The set $\mathcal{L}_+(\mathcal{H})$ of positive operators is a convex cone in this space. The set of strictly positive operators is denoted by $\mathcal{L}_{++}(\mathcal{H})$. It is an open set in $\mathcal{L}_{s.a.}(\mathcal{H})$ and is a convex cone. If f is a map of $\mathcal{L}_{s.a.}(\mathcal{H})$ into itself, we say f is *convex* if

$$f((1-\alpha)A + \alpha B) \le (1-\alpha)f(A) + \alpha f(B) \tag{1.30}$$

for all $A, B \in \mathcal{L}_{s.a.}(\mathcal{H})$ and for $0 \le \alpha \le 1$. If f is continuous, then f is convex if and only if

$$f\left(\frac{A+B}{2}\right) \le \frac{f(A)+f(B)}{2} \tag{1.31}$$

for all A, B.

We say f is *monotone* if $f(A) \ge f(B)$ whenever $A \ge B$.

The results on block matrices in Section 1.3 lead to easy proofs of the convexity and monotonicity of several functions. Here is a small sampler.

Let $A, B > O$. By Exercise 1.3.10

$$\begin{bmatrix} A & I \\ I & A^{-1} \end{bmatrix} \ge O \text{ and } \begin{bmatrix} B & I \\ I & B^{-1} \end{bmatrix} \ge O. \tag{1.32}$$

Hence,

$$\begin{bmatrix} A+B & 2I \\ 2I & A^{-1}+B^{-1} \end{bmatrix} \ge O.$$

By Theorem 1.3.3 this implies

$$A^{-1} + B^{-1} \geq 4(A+B)^{-1}$$

or

$$\left(\frac{A+B}{2}\right)^{-1} \leq \frac{A^{-1}+B^{-1}}{2}. \tag{1.33}$$

Thus the map $A \mapsto A^{-1}$ is convex on the set of positive matrices.

Taking the Schur product of the two block matrices in (1.32) we get

$$\begin{bmatrix} A \circ B & I \\ I & A^{-1} \circ B^{-1} \end{bmatrix} \geq O.$$

So, by Theorem 1.3.3

$$A \circ B \geq (A^{-1} \circ B^{-1})^{-1}. \tag{1.34}$$

The special choice $B = A^{-1}$ gives

$$A \circ A^{-1} \geq (A^{-1} \circ A)^{-1} = (A \circ A^{-1})^{-1}.$$

But a positive matrix is larger than its inverse if and only if it is larger than I. Thus we have the inequality

$$A \circ A^{-1} \geq I \tag{1.35}$$

known as *Fiedler's inequality.*

1.5.1 Exercise

Use Theorem 1.3.3 to show that the map $(B, X) \mapsto XB^{-1}X^*$ from $\mathcal{L}_{++}(\mathcal{H}) \times \mathcal{L}(\mathcal{H})$ into $\mathcal{L}_{+}(\mathcal{H})$ is jointly convex, i.e.,

$$\left(\frac{X_1+X_2}{2}\right)\left(\frac{B_1+B_2}{2}\right)^{-1}\left(\frac{X_1+X_2}{2}\right)^* \leq \frac{X_1B_1^{-1}X_1^* + X_2B_2^{-1}X_2^*}{2}.$$

In particular, this implies that the map $B \mapsto B^{-1}$ on $\mathcal{L}_{++}(\mathcal{H})$ is convex (a fact we have proved earlier), and that the map $X \mapsto X^2$ on $\mathcal{L}_{s.a.}(\mathcal{H})$ is convex. The latter statement can be proved directly: for all Hermitian matrices A and B we have the inequality

$$\left(\frac{A+B}{2}\right)^2 \leq \frac{A^2+B^2}{2}.$$

1.5.2 Exercise

Show that the map $X \mapsto X^3$ is not convex on 2×2 positive matrices.

1.5.3 Corollary

The map $(A, B, X) \mapsto A - XB^{-1}X^*$ *is jointly concave on* $\mathcal{L}_+(\mathcal{H}) \times \mathcal{L}_{++}(\mathcal{H}) \times \mathcal{L}(\mathcal{H})$*. It is monotone increasing in the variables* A, B*.*

In particular, the map $B \mapsto -B^{-1}$ on $\mathcal{L}_{++}(\mathcal{H})$ is monotone (a fact we proved earlier in Exercise 1.2.12).

1.5.4 Proposition

Let $B > O$ *and let* X *be any matrix. Then*

$$XB^{-1}X^* = \min \left\{ A : \begin{bmatrix} A & X \\ X^* & B \end{bmatrix} \geq O \right\}. \tag{1.36}$$

Proof. This follows immediately from Theorem 1.3.3. ■

1.5.5 Corollary

Let A, B *be positive matrices and* X *any matrix. Then*

$$A - XB^{-1}X^* = \max \left\{ Y : \begin{bmatrix} A & X \\ X^* & B \end{bmatrix} \geq \begin{bmatrix} Y & O \\ O & O \end{bmatrix} \right\}. \tag{1.37}$$

Proof. Use Proposition 1.5.4. ■

Extremal representations such as (1.36) and (1.37) are often used to derive matrix inequalities. Most often these are statements about convexity of certain maps. Corollary 1.5.5, for example, gives useful information about the Schur complement, a concept much used in matrix theory and in statistics.

Given a block matrix $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ the *Schur complement* of A_{22} in A is the matrix

$$\widetilde{A}_{22} = A_{11} - A_{12}A_{22}^{-1}A_{21}. \tag{1.38}$$

The Schur complement of A_{11} is the matrix obtained by interchanging the indices 1 and 2 in this definition. Two reasons for interest in this object are given below.

1.5.6 Exercise

Show that

$$\det A \;=\; \det \widetilde{A}_{22}\;\det A_{22}.$$

1.5.7 Exercise

If A is invertible, then the top left corner of the block matrix A^{-1} is $(\widetilde{A}_{11})^{-1}$; i.e.,

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} \;=\; \begin{bmatrix} (\widetilde{A}_{11})^{-1} & * \\ * & * \end{bmatrix}.$$

Corollary 1.5.3 says that on the set of positive matrices (with a block decomposition) the Schur complement is a concave function. Let us make this more precise. Let $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$ be an orthogonal decomposition of $\mathcal{H}$. Each operator A on $\mathcal{H}$ can be written as $A = \left[\begin{smallmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{smallmatrix}\right]$ with respect to this decomposition. Let $P(A) = \tilde{A}_{22}$. Then for all strictly positive operators A and B we have

$$P(A+B) \;\geq\; P(A) + P(B).$$

The map $A \mapsto P(A)$ is positively homogenous; i.e., $P(\alpha A) = \alpha P(A)$ for all positive numbers α. It is also monotone in A.

We have seen that while the function $f(A) = A^2$ is convex on the space of positive matrices, the function $f(A) = A^3$ is not; and while the function $f(A) = A^{1/2}$ is monotone on the set of positive matrices, the function $f(A) = A^2$ is not. Thus the following theorems are interesting.

1.5.8 Theorem

The function $f(A) = A^r$ is convex on $\mathcal{L}_+(\mathcal{H})$ for $1 \leq r \leq 2$.

Proof. We give a proof that uses Exercise 1.5.1 and a useful integral representation of A^r. For $t > 0$ and $0 < r < 1$, we have (from one of the integrals calculated via contour integration in complex analysis)

$$t^r = \frac{\sin r\pi}{\pi} \int_0^\infty \frac{t}{\lambda + t} \lambda^{r-1} d\lambda \; .$$

The crucial feature of this formula that will be exploited is that we can represent t^r as

$$t^r = \int_0^\infty \frac{t}{\lambda + t}\, d\mu(\lambda), \quad 0 < r < 1, \tag{1.39}$$

where μ is a positive measure on $(0, \infty)$. Multiplying both sides by t, we have

$$t^r = \int_0^\infty \frac{t^2}{\lambda + t}\, d\mu(\lambda), \qquad 1 < r < 2.$$

Thus for positive operators A, and for $1 < r < 2$,

$$\begin{aligned} A^r &= \int_0^\infty A^2(\lambda + A)^{-1} d\mu(\lambda) \\ &= \int_0^\infty A(\lambda + A)^{-1} A \;\; d\mu(\lambda). \end{aligned}$$

By Exercise 1.5.1 the integrand is a convex function of A for each $\lambda > 0$. So the integral is also convex. ■

1.5.9 Theorem

The function $f(A) = A^r$ is monotone on $\mathcal{L}_+(\mathcal{H})$ for $0 \le r \le 1$.

Proof. For each $\lambda > 0$ we have $A(\lambda + A)^{-1} = (\lambda A^{-1} + I)^{-1}$. Since the map $A \mapsto A^{-1}$ is monotonically decreasing (Exercise 1.2.12), the function $f_\lambda(A) = A(\lambda + A)^{-1}$ is monotonically increasing for each $\lambda > 0$. Now use the integral representation (1.39) as in the preceding proof. ■

1.5.10 Exercise

Show that the function $f(A) = A^r$ is convex on $\mathcal{L}_{++}(\mathcal{H})$ for $-1 \le r \le 0$. [Hint: For $-1 < r < 0$ we have

$$t^r = \int_0^\infty \frac{1}{\lambda + t} d\mu(\lambda).$$

Use the convexity of the map $f(A) = A^{-1}$.]

1.6 SUPPLEMENTARY RESULTS AND EXERCISES

Let A and B be Hermitian matrices. If there exists an invertible matrix X such that X^*AX and X^*BX are diagonal, we say that A and B are *simultaneously congruent to diagonal matrices* (or A and B are *simultaneously diagonalizable by a congruence*). If X can be chosen to be unitary, we say A and B are *simultaneously diagonalizable by a unitary conjugation.*

Two Hermitian matrices are simultaneously diagonalizable by a unitary conjugation if and only if they commute. Simultaneous congruence to diagonal matrices can be achieved under less restrictive conditions.

1.6.1 Exercise

Let A be a strictly positive and B a Hermitian matrix. Then A and B are simultaneously congruent to diagonal matrices. [Hint: A is congruent to the identity matrix.]

Simultaneous diagonalization of *three* matrices by congruence, however, again demands severe restrictions. Consider three strictly positive matrices I, A_1 and A_2. Suppose X is an invertible matrix such that X^*X is diagonal. Then $X = U\Lambda$ where U is unitary and Λ is diagonal and invertible. It is easy to see that for such an X, X^*A_1X and X^*A_2X both are diagonal if and only if A_1 and A_2 commute.

If A and B are Hermitian matrices, then the inequality $AB + BA \le A^2 + B^2$ is always true. It follows that if A and B are positive, then

$$\left(\frac{A^{1/2} + B^{1/2}}{2}\right)^2 \le \frac{A+B}{2}.$$

Using the monotonicity of the square root function we see that

$$\frac{A^{1/2}+B^{1/2}}{2} \leq \left(\frac{A+B}{2}\right)^{1/2}.$$

In other words the function $f(A) = A^{1/2}$ is *concave* on the set $\mathcal{L}_+(\mathcal{H})$.

More generally, it can be proved that $f(A) = A^r$ is concave on $\mathcal{L}_+(\mathcal{H})$ for $0 \leq r \leq 1$. See Theorem 4.2.3.

It is known that the map $f(A) = A^r$ on positive matrices is monotone if and only if $0 \leq r \leq 1$, and convex if and only if $r \in [-1, 0] \cup [1, 2]$. A detailed discussion of matrix monotonicity and convexity may be found in MA, Chapter V. Some of the proofs given here are different. We return to these questions in later chapters.

Given a matrix A let $A^{(m)}$ be the m-fold Schur product $A \circ A \circ \cdots \circ A$. If A is positive semidefinite, then so is $A^{(m)}$. Suppose all the entries of A are nonnegative real numbers a_{ij}. In this case we say that A is *entrywise positive,* and for each $r > 0$ we define $A^{(r)}$ as the matrix whose entries are a_{ij}^r. If A is entrywise positive and positive semidefinite, then $A^{(r)}$ is not always positive semidefinite. For example, let

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

and consider $A^{(r)}$ for $0 < r < 1$.

An entrywise positive matrix is said to be *infinitely divisible* if the matrix $A^{(r)}$ is positive semidefinite for all $r > 0$.

1.6.2 Exercise

Show that if A is an entrywise positive matrix and $A^{(1/m)}$ is positive semidefinite for all natural numbers m, then A is infinitely divisible.

In the following two exercises we outline proofs of the fact that the Cauchy matrix (1.5) is infinitely divisible.

1.6.3 Exercise

Let $\lambda_1, \ldots, \lambda_m$ be positive numbers and let $\varepsilon > 0$ be any lower bound for them. For $r > 0$, let $C_\varepsilon^{(r)}$ be the matrix whose i, j entries are

$$\frac{1}{(\lambda_i + \lambda_j - \varepsilon)^r}.$$

Write these numbers as

$$\left(\frac{\varepsilon}{\lambda_i\lambda_j}\right)^r \left(\frac{1}{1-\frac{(\lambda_i-\varepsilon)(\lambda_j-\varepsilon)}{\lambda_i\lambda_j}}\right)^r.$$

Use this to show that $C_\varepsilon^{(r)}$ is congruent to the matrix whose i, j entries are

$$\left(\frac{1}{1-\frac{(\lambda_i-\varepsilon)(\lambda_j-\varepsilon)}{\lambda_i\lambda_j}}\right)^r = \sum_{n=0}^{\infty} a_n \left(\frac{(\lambda_i-\varepsilon)(\lambda_j-\varepsilon)}{\lambda_i\lambda_j}\right)^n.$$

The coefficients a_n are the numbers occurring in the binomial expansion

$$\left(\frac{1}{1-x}\right)^r = \sum_{n=0}^{\infty} a_n x^n, \quad |x| < 1,$$

and are positive. The matrix with entries

$$\frac{(\lambda_i-\varepsilon)(\lambda_j-\varepsilon)}{\lambda_i\lambda_j}$$

is congruent to the matrix with all its entries equal to 1. So, it is positive semidefinite. It follows that $C_\varepsilon^{(r)}$ is positive semidefinite for all $\varepsilon > 0$. Let $\varepsilon \downarrow 0$ and conclude that the Cauchy matrix is infinitely divisible.

1.6.4 Exercise

The gamma function for $x > 0$ is defined by the formula

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt.$$

Show that for every $r > 0$ we have

$$\frac{1}{(\lambda_i+\lambda_j)^r} = \frac{1}{\Gamma(r)} \int_0^\infty e^{-t(\lambda_i+\lambda_j)} t^{r-1} dt.$$

This shows that the matrix $\left[\left[\frac{1}{(\lambda_i+\lambda_j)^r}\right]\right]$ is a Gram matrix, and gives another proof of the infinite divisibility of the Cauchy matrix.

1.6.5 Exercise

Let $\lambda_1, \ldots, \lambda_n$ be positive numbers and let Z be the $n \times n$ matrix with entries

$$z_{ij} = \frac{1}{\lambda_i^2 + \lambda_j^2 + t\lambda_i\lambda_j},$$

where $t > -2$. Show that for all $t \in (-2, 2]$ this matrix is infinitely divisible. [Hint: Use the expansion

$$z_{ij}^r = \frac{1}{(\lambda_i + \lambda_j)^{2r}} \sum_{m=0}^{\infty} a_m (2-t)^m \frac{\lambda_i^m \lambda_j^m}{(\lambda_i + \lambda_j)^{2m}}.]$$

Let $n = 2$. Show that the matrix $Z^{(r)}$ in this case is positive semidefinite for $t \in (-2, \infty)$ and $r > 0$.

Let $(\lambda_1, \lambda_2, \lambda_3) = (1, 2, 3)$ and $t = 10$. Show that with this choice the 3×3 matrix Z is not positive semidefinite.

In Chapter 5 we will study this example again and show that the condition $t \in (-2, 2]$ is necessary to ensure that the matrix Z defined above is positive semidefinite for all n and all positive numbers $\lambda_1, \ldots, \lambda_n$.

If $A = [[a_{ij}]]$ is a positive matrix, then so is its complex conjugate $\overline{A} = [[\overline{a}_{ij}]]$. The Schur product of these two matrices $[[\ |a_{ij}|^2]]$ is positive, as are all the matrices $[[\ |a_{ij}|^{2k}]]$, $k = 0, 1, 2, \ldots$.

1.6.6 Exercise

(i) Let $n \leq 3$ and let $[[a_{ij}]]$ be an $n \times n$ positive matrix. Show that the matrix $[[\ |a_{ij}|\ ]]$ is positive.

(ii) Let

$$A = \begin{bmatrix} 1 & \frac{1}{\sqrt{2}} & 0 & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 1 & \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{1}{\sqrt{2}} & 1 & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 1 \end{bmatrix}.$$

Show that A is positive but $[[\ |a_{ij}|\ ]]$ is not.

Let $\varphi : \mathbb{C} \to \mathbb{C}$ be a function satisfying the following property: whenever A is a positive matrix (of any size), then $[[\varphi(a_{ij})]]$ is positive.

It is known that such a function has a representation as a series

$$\varphi(z) = \sum_{k,l=0}^{\infty} b_{kl}\, z^k\, \overline{z}^l, \tag{1.40}$$

that converges for all z, and in which all coefficients b_{kl} are nonnegative.

From this it follows that if p is a positive real number but not an even integer, then there exists a positive matrix A (of some size n depending on p) such that $[[\ |a_{ij}|^p\ ]]$ is not positive.

Since $\|A\| = \|\,|A|\,\|$ for all operators A, the triangle inequality may be expressed also as

$$\|A + B\| \leq \|\,|A|\,\| + \|\,|B|\,\| \quad \text{for all } A, B \in \mathcal{L}(\mathcal{H}). \tag{1.41}$$

If both A and B are normal, this can be improved. Using Corollary 1.3.8 we see that in this case

$$\begin{bmatrix} |A| + |B| & A^* + B^* \\ A + B & |A| + |B| \end{bmatrix} \geq O.$$

Then using Proposition 1.3.2 we obtain

$$\|A + B\| \leq \|\,|A| + |B|\,\| \text{ for } A, B \text{ normal}. \tag{1.42}$$

This inequality is stronger than (1.41). It is not true for all A and B, as may be seen from the example

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

The inequality (1.42) has an interesting application in the proof of Theorem 1.6.8 below.

1.6.7 Exercise

Let A and B be any two operators, and for a given positive integer m let $\omega = e^{2\pi i/m}$. Prove the identity

$$A^m + B^m = \frac{(A + B)^m + (A + \omega B)^m + \cdots + (A + \omega^{m-1} B)}{m}. \tag{1.43}$$

1.6.8 Theorem

Let A and B be positive operators. Then

$$\|A^m + B^m\| \le \|(A+B)^m\| \quad \text{for } m = 1, 2, \ldots. \tag{1.44}$$

Proof. Using the identity (1.43) we get

$$\begin{aligned} \|A^m + B^m\| &\le \frac{1}{m}\sum_{j=0}^{m-1} \|(A+\omega^j B)^m\| \\ &\le \frac{1}{m}\sum_{j=0}^{m-1} \|A+\omega^j B\|^m. \end{aligned} \tag{1.45}$$

For each complex number z, the operator zB is normal. So from (1.42) we get

$$\|A + zB\| \le \|A + \ |z|\ B\|.$$

This shows that each of the summands in the sum on the right-hand side of (1.45) is bounded by $\|A+B\|^m$. Since $A+B$ is positive, $\|A+B\|^m = \|(A+B)^m\|$. This proves the theorem. ■

The next theorem is more general.

1.6.9 Theorem

Let A and B be positive operators. Then

$$\|A^r + B^r\| \le \|(A+B)^r\| \quad \text{for } 1 \le r < \infty, \tag{1.46}$$

$$\|A^r + B^r\| \ge \|(A+B)^r\| \quad \text{for } 0 \le r \le 1. \tag{1.47}$$

Proof. Let m be any positive integer and let Ω_m be the set of all real numbers r in the interval $[1, m]$ for which the inequality (1.46) is true. We will show that Ω_m is a convex set. Since 1 and m belong to Ω_m, this will prove the inequality (1.46).

Suppose r and s are two points in Ω_m and let $t = (r+s)/2$. Then

$$\begin{bmatrix} A^t + B^t & O \\ O & O \end{bmatrix} = \begin{bmatrix} A^{r/2} & B^{r/2} \\ O & O \end{bmatrix} \begin{bmatrix} A^{s/2} & O \\ B^{s/2} & O \end{bmatrix}.$$

Hence

$$\|A^t + B^t\| \le \left\| \begin{bmatrix} A^{r/2} & B^{r/2} \\ O & O \end{bmatrix} \right\| \left\| \begin{bmatrix} A^{s/2} & O \\ B^{s/2} & O \end{bmatrix} \right\|.$$

Since $\|X\| = \|X^*X\|^{1/2} = \|XX^*\|^{1/2}$ for all X, this gives

$$\|A^t + B^t\| \leq \|A^r + B^r\|^{1/2} \ \|A^s + B^s\|^{1/2}.$$

We have assumed r and s are in Ω_m. So, we have

$$\begin{aligned}\|A^t + B^t\| &\leq \|(A+B)^r\|^{1/2} \ \|(A+B)^s\|^{1/2} \\ &= \|A+B\|^{r/2}\|A+B\|^{s/2} \\ &= \|A+B\|^t = \|(A+B)^t\|.\end{aligned}$$

This shows that $t \in \Omega_m$, and the inequality (1.46) is proved.

Let $0 < r \leq 1$. Then from (1.46) we have

$$\|A^{1/r} + B^{1/r}\| \leq \|(A+B)^{1/r}\| = \|A+B\|^{1/r}.$$

Replacing A and B by A^r and B^r, we obtain the inequality (1.47). ■

We have seen that $AB+BA$ need not be positive when A and B are positive. Hence we do not always have $A^2 + B^2 \leq (A+B)^2$. Theorem 1.6.8 shows that we do have the weaker assertion

$$\lambda_1^{\downarrow}(A^2 + B^2) \leq \lambda_1^{\downarrow}(A+B)^2.$$

1.6.10 Exercise

Use the example

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

to see that the inequality

$$\lambda_2^{\downarrow}(A^2 + B^2) \leq \lambda_2^{\downarrow}(A+B)^2$$

is not always true.

1.7 NOTES AND REFERENCES

Chapter 7 of the well known book *Matrix Analysis* by R. A. Horn and C. R. Johnson, Cambridge University Press, 1985, is an excellent source of information about the basic properties of positive definite matrices. See also Chapter 6 of F. Zhang, *Matrix Theory: Basic Results and Techniques,* Springer, 1999. The reader interested in numerical analysis should see Chapter 10 of N. J. Higham, *Accuracy and*

Stability of Numerical Algorithms, Second Edition, SIAM 2002, and Chapter 5 of G. H. Golub and C. F. Van Loan, *Matrix Computations,* Third Edition, Johns Hopkins University Press, 1996.

The matrix in (1.5) is a special case of the more general matrix C whose entries are

$$c_{ij} = \frac{1}{\lambda_i + \mu_j},$$

where $(\lambda_1, \ldots, \lambda_m)$ and $(\mu_1, \ldots, \mu_m)$ are any two real m-tuples. In 1841, Cauchy gave a formula for the determinant of this matrix:

$$\det C = \frac{\prod\limits_{1 \le i < j \le m} (\lambda_j - \lambda_i)(\mu_j - \mu_i)}{\prod\limits_{1 \le i,\, j \le m} (\lambda_i + \mu_j)}.$$

From this it follows that the matrix in (1.5) is positive. The Hilbert matrix H with entries

$$h_{ij} = \frac{1}{i + j - 1}$$

is a special kind of Cauchy matrix. Hilbert showed that the infinite matrix H defines a bounded operator on the space l_2 and $\|H\| < 2\pi$. The best value π for $\|H\|$ was obtained by I. Schur, *Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen,* J. Reine Angew. Math., 140 (1911). This is now called Hilbert's inequality. See Chapter IX of G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities,* Second Edition, Cambridge University Press, 1952, for different proofs and interesting extensions. Chapter 28 of Higham's book, cited earlier, describes the interest that the Hilbert and Cauchy matrices have for the numerical analyst.

Sylvester's law of inertia is an important fact in the reduction of a real quadratic form to a sum of squares. While such a reduction may be achieved in many different ways, the law says that the number of positive and the number of negative squares are always the same in any such reduced representation. See Chapter X of F. Gantmacher, *Matrix Theory,* Chelsea, 1977. The law has special interest in the stability theory of differential equations where many problems depend on information on the location of the eigenvalues of matrices in the left half-plane. A discussion of these matters, and of the Lyapunov and Stein equations, may be found in P. Lancaster and M. Tismenetsky, *The Theory of Matrices,* Second Edition, Academic Press, 1985.

The Descartes rule of signs (in a slightly refined form) says that if all the roots of a polynomial $f(x)$ are real, and the constant term is nonzero, then the number k_1 of positive roots of the polynomial is equal to the number s_1 of variations in sign in the sequence of its coefficients, and the number k_2 of negative roots is equal to the number s_2 of variations in sign in the sequence of coefficients of the polynomial $f(-x)$. See, for example, A. Kurosh, *Higher Algebra,* Mir Publishers, 1972.

The symmetrized product (1.12), divided by 2, is also called the *Jordan product.* In Chapter 10 of P. Lax, *Linear Algebra,* John Wiley, 1997, the reader will find different (and somewhat longer) proofs of Propositions 1.2.6 and 1.2.9. Proposition 1.2.9 and Theorem 1.5.9 are a small part of Loewner's theory of operator monotone functions. An exposition of the full theory is given in Chapter V of R. Bhatia, *Matrix Analysis,* Springer, 1997 (abbreviated to MA in our discussion).

The equation $XAX = B$ in Exercise 1.2.13 is a very special kind of Riccati equation, the general form of such an equation being

$$XAX - XC - C^*X = B.$$

Such equations arise in problems of control theory, and have been studied extensively. See, for example, P. Lancaster and L. Rodman, *The Algebraic Riccati Equation,* Oxford University Press, 1995.

Proposition 1.3.1 makes connection between an algebraic property–positivity, and a metric property—contractivity. The technique of studying properties of a matrix by embedding it in a larger matrix is known as "dilation theory" and has proved to be of great value. An excellent demonstration of the power of such methods is given in the two books by V. Paulsen, *Completely Bounded Maps and Dilations,* Longman, 1986 and *Completely Bounded Maps and Operator Algebras,* Cambridge University Press, 2002. Theorem 1.3.3 has been used to great effect by several authors. The idea behind this 2×2 matrix calculation can be traced back to I. Schur, *Über Potenzreihen die im Innern des Einheitskreises beschränkt sind* [I], J. Reine Angew. Math., 147 (1917) 205–232, where the determinantal identity of Exercise 1.5.6 occurs. Using the idea in our first proof of Theorem 1.3.3 it is easy to deduce the following. If $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ is a Hermitian matrix and $\widetilde{A}_{22}$ is defined by (1.38), then we have the relation

$$\text{In}\,(A) = \text{In}(A_{22}) + \text{In}(\widetilde{A}_{22})$$

between inertias. This fact was proved by E. Haynsworth, *Determination of the inertia of a partitioned Hermitian matrix,* Linear Algebra Appl., 1 (1968) 73–81. The term *Schur complement* was introduced by Haynsworth. The argument of Theorem 1.3.3 is a special case of this additivity of inertias.

The Schur complement is important in matrix calculations arising in several areas like statistics and electrical engineering. The recent book, *The Schur Complement and Its Applications,* edited by F. Zhang, Springer, 2005, contains several expository articles and an exhaustive bibliography that includes references to earlier expositions. The Schur complement is used in quantum mechanics as the *decimation map* or the *Feshbach map.* Here the Hamiltonian is a Hermitian matrix written in block form as

$$H = \begin{bmatrix} A & B \\ C & D \end{bmatrix}.$$

The block A corresponds to low energy states of the system without interaction. The *decimated Hamiltonian* is the matrix $A - BD^{-1}C$. See W. G. Faris, *Review of S. J. Gustafson and I. M. Sigal, Mathematical Concepts of Quantum Mechanics,* SIAM Rev., 47 (2005) 379–380.

Perhaps the best-known theorem about the product $A \circ B$ is that it is positive when A and B are positive. This, together with other results like Theorem 1.4.1, was proved by I. Schur in his 1911 paper cited earlier.

For this reason, this product has been called the Schur product. *Hadamard product* is another name for it. The entertaining and informative article R. Horn, *The Hadamard product,* in *Matrix Theory and Applications,* C. R. Johnson, ed., American Math. Society, 1990, contains a wealth of historical and other detail. Chapter 5 of R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis,* Cambridge University Press, 1991 is devoted to this topic. Many recent theorems, especially inequalities involving the Schur product, are summarised in the report T. Ando, *Operator-Theoretic Methods for Matrix Inequalities,* Sapporo, 1998.

Monotone and convex functions of self-adjoint operators have been studied extensively since the appearance of the pioneering paper by K. Löwner, *Über monotone Matrixfunctionen,* Math. Z., 38 (1934) 177–216. See Chapter V of MA for an introduction to this topic. The elegant and effective use of block matrices in this context is mainly due to T. Ando, *Topics on Operator Inequalities,* Hokkaido University, Sapporo, 1978, and *Concavity of certain maps on positive definite matrices and applications to Hadamard products,* Linear Algebra Appl. 26 (1979) 203–241.

The representation (1.40) for functions that preserve positivity (in the sense described) was established in C. S. Herz, *Fonctions opérant* *sur les fonctions définies-positives,* Ann. Inst. Fourier (Grenoble), 13 (1963) 161–180. A real variable version of this was noted earlier by I. J. Schoenberg.

Theorems 1.6.8 and 1.6.9 were proved in R. Bhatia and F. Kittaneh, *Norm inequalities for positive operators,* Lett. Math. Phys. 43 (1998) 225–231. A much more general result, conjectured in this paper and proved by T. Ando and X. Zhan, *Norm inequalities related to operator monotone functions,* Math. Ann., 315 (1999) 771–780, says that for every nonnegative operator monotone function f on $[0, \infty)$ we have $\|f(A+B)\| \leq \|f(A)+f(B)\|$ for all positive matrices A and B. Likewise, if g is a nonnegative increasing function on $[0, \infty)$ such that $g(0) = 0$, $g(\infty) = \infty$, and the inverse function of g is operator monotone, then $\|g(A)+g(B)\| \leq \|g(A+B)\|$. This includes Theorem 1.6.9 as a special case. Further, these inequalities are valid for a large class of norms called unitarily invariant norms. (Operator monotone functions and unitarily invariant norms are defined in Section 2.7.) It may be of interest to mention here that, with the notations of this paragraph, we have also the inequalities $\|f(A) - f(B)\| \leq \|f(|A-B|)\|$, and $\|g(|A-B|)\| \leq \|g(A) - g(B)\|$. See Theorems X.1.3 and X.1.6 in MA. More results along these lines can be found in X. Zhan, *Matrix Inequalities,* Lecture Notes in Mathematics, Vol. 1790, Springer, 2002.

Many important and fundamental theorems of the rapidly developing subject of quantum information theory are phrased as inequalities for positive matrices (often block matrices). One popular book on this subject is M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information,* Cambridge University Press, 2002.

Chapter Two

Positive Linear Maps

In this chapter we study linear maps on spaces of matrices. We use the symbol Φ for a linear map from $\mathbb{M}_n$ into $\mathbb{M}_k$. When $k = 1$ such a map is called a linear functional, and we use the lower-case symbol φ for it. The norm of Φ is

$$\|\Phi\| = \sup_{\|A\|=1} \|\Phi(A)\| = \sup_{\|A\|\leq 1} \|\Phi(A)\|.$$

In general, it is not easy to calculate this. One of the principal results of this chapter is that if Φ carries positive elements of $\mathbb{M}_n$ to positive elements of $\mathbb{M}_k$, then $\|\Phi\| = \|\Phi(I)\|$.

2.1 REPRESENTATIONS

The interplay between algebraic properties of linear maps Φ and their metric properties is best illustrated by considering *representations* of $\mathbb{M}_n$ in $\mathbb{M}_k$. These are linear maps that

(i) preserve products; i.e., $\Phi(AB) = \Phi(A)\Phi(B)$;

(ii) preserve adjoints; i.e., $\Phi(A^*) = \Phi(A)^*$;

(iii) preserve the identity; i.e., $\Phi(I) = I$.

Let $\sigma(A)$ denote the spectrum of A, and $\operatorname{spr}(A)$ its spectral radius.

2.1.1 Exercise

If Φ has properties (i) and (iii), then

$$\sigma(\Phi(A)) \subset \sigma(A). \tag{2.1}$$

Hence

$$\operatorname{spr}(\Phi(A)) \le \operatorname{spr}(A). \tag{2.2}$$

Our norm $\|\cdot\|$ has two special properties related to the $*$ operation: $\|A\|^2 = \|A^*A\|$; and $\|A\| = \operatorname{spr}(A)$ if A is Hermitian. So, if Φ is a representation we have

$$\begin{aligned}\|\Phi(A)\|^2 = \|\Phi(A)^*\Phi(A)\| &= \|\Phi(A^*A)\| = \operatorname{spr}(\Phi(A^*A))\\ \le \operatorname{spr}(A^*A) &= \|A^*A\| = \|A\|^2.\end{aligned}$$

Thus $\|\Phi(A)\| \le \|A\|$ for all A. Since $\Phi(I) = I$, we have $\|\Phi\| = 1$.

We have shown that *every representation has norm one.*

How does one get representations? For each unitary element of $\mathbb{M}_n$, $\Phi(A) = U^*AU$ is a representation. Direct sums of such maps are representations; i.e., if $U_1, \ldots, U_r$ are $n \times n$ unitary matrices, then $\Phi(A) = U_1^*AU_1 \oplus \cdots \oplus U_r^*AU_r$ is a representation.

Choosing $U_j = I$, $1 \le j \le r$, we get the representation $\Phi(A) = I_r \otimes A$. The operator $A \otimes I_r$ is unitarily equivalent to $I_r \otimes A$, and $\Phi(A) = A \otimes I_r$ is another representation.

2.1.2 Exercise

All representations of $\mathbb{M}_n$ are obtained by composing unitary conjugations and tensor products with I_r, $r = 1, 2, \ldots$. Thus we have exhausted the family of representations by the examples we saw above. [Hint: A representation carries orthogonal projections to orthogonal projections, unitaries to unitaries, and preserves unitary conjugation.]

Thus the fact that $\|\Phi\| = 1$ for every representation Φ is not too impressive; we do know $\|I \otimes A\| = \|A\|$ and $\|U^*AU\| = \|A\|$.

We will see how we can replace the multiplicativity condition (i) by less restrictive conditions and get a richer theory.

2.2 POSITIVE MAPS

A linear map $\Phi : \mathbb{M}_n \to \mathbb{M}_k$ is called *positive* if $\Phi(A) \ge O$ whenever $A \ge O$. It is said to be *unital* if $\Phi(I) = I$. We will say Φ is strictly positive if $\Phi(A) > O$ whenever $A > O$. It is easy to see that a positive linear map Φ is strictly positive if and only if $\Phi(I) > O$.

2.2.1 Examples

(i) $\varphi(A) = \text{tr}A$ is a positive linear functional; $\varphi(A) = \frac{1}{n}\text{tr}A$ is positive and unital.

(ii) Every linear functional on $\mathbb{M}_n$ has the form $\varphi(A) = \text{tr}AX$ for some $X \in \mathbb{M}_n$. It is easy to see that φ is positive if and only if X is a positive matrix; φ is unital if $\text{tr}X = 1$. (Positive matrices of trace one are called *density matrices* in the physics literature.)

(iii) Let $\varphi(A) = \sum\limits_{i,j} a_{ij}$, the sum of all entries of A. If e is the vector with all of its entries equal to one, and $E = ee^*$, the matrix with all entries equal to one, then

$$\varphi(A) = \langle e, Ae \rangle = \text{ tr } AE.$$

Thus φ is a positive linear functional.

(iv) The map $\Phi(A) = \frac{\text{tr}A}{n} I$ is a positive map of $\mathbb{M}_n$ into itself. (Its range consists of scalar matrices.)

(v) Let A^{tr} denote the transpose of A. Then the map $\Phi(A) = A^{\text{tr}}$ is positive.

(vi) Let X be an $n \times k$ matrix. Then $\Phi(A) = X^*AX$ is a positive map from $\mathbb{M}_n$ into $\mathbb{M}_k$.

(vii) A special case of this is the *compression* map that takes an $n \times n$ matrix to a $k \times k$ block in its top left corner.

(viii) Let $P_1, \dots, P_r$ be mutually orthogonal projections with $P_1 \oplus \cdots \oplus P_r = I$. The operator $\Phi(A) = \sum P_j A P_j$ is called a *pinching* of A. In an appropriate coordinate system this can be described as

$$A = \begin{bmatrix} A_{11} & \dots & A_{1r} \\ A_{21} & \dots & A_{2r} \\ \cdot & \dots & \cdot \\ A_{r1} & \dots & A_{rr} \end{bmatrix}, \quad \mathcal{C}(A) = \begin{bmatrix} A_{11} & & & \\ & A_{22} & & \\ & & \ddots & \\ & & & A_{rr} \end{bmatrix}.$$

Every pinching is positive. A special case of this is $r = n$ and each P_j is the projection onto the linear span of the basis vector e_j. Then $\mathcal{C}(A)$ is the diagonal part of A.

(ix) Let B be any positive matrix. Then the map $\Phi(A) = A \otimes B$ is positive. So is the map $\Phi(A) = A \circ B$.

(x) Let A be a matrix whose spectrum is contained in the open right half plane. Let $L_A(X) = A^*X + XA$. The operator L_A on $\mathbb{M}_n$ is invertible and its inverse L_A^{-1} is a positive linear map. (See the discussion in Exercise 1.2.10.)

(xi) Any positive linear combination of positive maps is positive. Any convex combination of positive unital maps is positive and unital.

It is instructive to think of positive maps as noncommutative (matrix) averaging operations. Let $\mathrm{C}(X)$ be the space of continuous functions on a compact metric space. Let φ be a linear functional on $\mathrm{C}(X)$. By the Riesz representation theorem, there exists a signed measure μ on X such that

$$\varphi(f) = \int f d\mu. \tag{2.3}$$

The linear functional φ is called positive if $\varphi(f) \geq 0$ for every (pointwise) nonnegative function f. For such a φ, the measure μ representing it is a positive measure. If φ maps the function $f \equiv 1$ to the number 1, then φ is said to be unital, and then μ is a probability measure. The integral (2.3) is then written as

$$\varphi(f) = Ef, \tag{2.4}$$

and called the *expectation* of f. Every positive, unital, linear functional on $\mathrm{C}(X)$ is an expectation (with respect to a probability measure μ). A positive, unital, linear map Φ may thus be thought of as a noncommutative analogue of an expectation map.

2.3 SOME BASIC PROPERTIES OF POSITIVE MAPS

We prove three theorems due to Kadison, Choi, and Russo and Dye. Our proofs use 2×2 block matrix arguments.

2.3.1 Lemma

Every positive linear map is adjoint-preserving; i.e., $\Phi(T^*) = \Phi(T)^*$ for all T.

Proof. First we show that $\Phi(A)$ is Hermitian if A is Hermitian. Every Hermitian matrix A has a *Jordan decomposition*

$$A = A_+ - A_- \qquad \text{where } A_\pm \geq O.$$

So,

$$\Phi(A) = \Phi(A_+) - \Phi(A_-)$$

is the difference of two positive matrices, and is therefore Hermitian. Every matrix T has a *Cartesian decomposition*

$$T = A + iB \qquad \text{where } A, B \text{ are Hermitian.}$$

So,

$$\Phi(T)^* = \Phi(A) - i\Phi(B) = \Phi(A - iB) = \Phi(T^*). \blacksquare$$

2.3.2 Theorem (Kadison's Inequality)

Let Φ be positive and unital. Then for every Hermitian A

$$\Phi(A)^2 \leq \Phi(A^2). \tag{2.5}$$

Proof. By the spectral theorem, $A = \sum \lambda_j P_j$, where λ_j are the eigenvalues of A and P_j the corresponding projections with $\sum P_j = I$. Then $A^2 = \sum \lambda_j^2 P_j$ and

$$\Phi(A) = \sum \lambda_j \Phi(P_j),\ \Phi(A^2) = \sum \lambda_j^2 \Phi(P_j),\ \sum \Phi(P_j) = I.$$

Since P_j are positive, so are $\Phi(P_j)$. Therefore,

$$\begin{bmatrix} \Phi(A^2) & \Phi(A) \\ \Phi(A) & I \end{bmatrix} = \sum \begin{bmatrix} \lambda_j^2 & \lambda_j \\ \lambda_j & 1 \end{bmatrix} \otimes \Phi(P_j).$$

Each summand in the last sum is positive and, hence, so is the sum. By Theorem 1.3.3, therefore,

$$\Phi(A^2) \geq \Phi(A)I^{-1}\Phi(A) = \Phi(A)^2. \blacksquare$$

2.3.3 Exercise

The inequality (2.5) may not be true if Φ is not unital.

Recall that for real functions we have $(Ef)^2 \leq Ef^2$. The inequality (2.5) is a noncommutative version of this. It should be pointed out that not all inequalities for expectations of real functions have such noncommutative counterparts. For example, we do have $(Ef)^4 \leq Ef^4$, but the analogous inequality $\Phi(A)^4 \leq \Phi(A^4)$ is not always true. To see this, let Φ be the compression map from $\mathbb{M}_3$ to $\mathbb{M}_2$, taking a 3×3 matrix to its top left 2×2 submatrix. Let

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

Then $\Phi(A)^4 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $\Phi(A^4) = \begin{bmatrix} 9 & 5 \\ 5 & 3 \end{bmatrix}$.

This difference can be attributed to the fact that while the function $f(t) = t^4$ is convex on the real line, the matrix function $f(A) = A^4$ is not convex on Hermitian matrices.

The following theorem due to Choi generalizes Kadison's inequality to normal operators.

2.3.4 Theorem (Choi)

Let Φ be positive and unital. Then for every normal matrix A

$$\Phi(A)\Phi(A^*) \leq \Phi(A^*A), \quad \Phi(A^*)\Phi(A) \leq \Phi(A^*A). \tag{2.6}$$

Proof. The proof is similar to the one for Theorem 2.3.2. We have

$$A = \sum \lambda_j P_j, \;\; A^* = \sum \overline{\lambda_j} P_j, \;\; A^*A = \sum |\lambda_j|^2 P_j.$$

So

$$\begin{bmatrix} \Phi(A^*A) & \Phi(A) \\ \Phi(A^*) & I \end{bmatrix} = \sum \begin{bmatrix} |\lambda_j|^2 & \lambda_j \\ \overline{\lambda_j} & 1 \end{bmatrix} \otimes \Phi(P_j)$$

is positive. ■

In Chapter 3, we will see that the condition that A be normal can be dropped if we impose a stronger condition (2-positivity) on Φ.

2.3.5 Exercise

If A is normal, then $\Phi(A)$ need not be normal. Thus the left-hand sides of the two inequalities (2.6) can be different.

2.3.6 Theorem (Choi's Inequality)

Let Φ be strictly positive and unital. Then for every strictly positive matrix A

$$\Phi(A)^{-1} \leq \Phi(A^{-1}). \tag{2.7}$$

Proof. The proof is again similar to that of Theorem 2.3.2. Now we have $A = \sum \lambda_j P_j$ with $\lambda_j > 0$. Then $A^{-1} = \sum \lambda_j^{-1} P_j$, and

$$\begin{bmatrix} \Phi(A^{-1}) & I \\ I & \Phi(A) \end{bmatrix} = \sum \begin{bmatrix} \lambda_j^{-1} & 1 \\ 1 & \lambda_j \end{bmatrix} \otimes \Phi(P_j)$$

is positive. Hence, by Theorem 1.3.3

$$\Phi(A^{-1}) \geq \Phi(A)^{-1}. \quad ■$$

2.3.7 Theorem (The Russo-Dye Theorem)

If Φ is positive and unital, then $\|\Phi\| = 1$.

Proof. We show first that $\|\Phi(U)\| \leq 1$ when U is unitary. In this case the eigenvalues λ_j are complex numbers of modulus one. So, from the spectral resolution $U = \sum \lambda_j P_j$, we get

$$\begin{bmatrix} I & \Phi(U) \\ \Phi(U)^* & I \end{bmatrix} = \sum \begin{bmatrix} 1 & \lambda_j \\ \overline{\lambda_j} & 1 \end{bmatrix} \otimes \Phi(P_j) \geq O.$$

Hence, by Proposition 1.3.1, $\|\Phi(U)\| \leq 1$. Now if A is any contraction, then we can write $A = \frac{1}{2}(U+V)$ where U, V are unitary. (Use the singular value decomposition of A and observe that if $0 \leq s \leq 1$, then we have $s = \frac{1}{2}(e^{i\theta} + e^{-i\theta})$ for some θ.) So

$$\|\Phi(A)\| = \frac{1}{2}\|\Phi(U+V)\| \leq \frac{1}{2}(\|\Phi(U)\| + \|\Phi(V)\|) \leq 1.$$

Thus $\|\Phi\| \leq 1$, and since Φ is unital $\|\Phi\| = 1$. ■

Second proof. Let $\|A\| \leq 1$. Then A has a *unitary dilation* $\hat{A}$

$$\hat{A} = \begin{bmatrix} A & -(I - AA^*)^{1/2} \\ (I - A^*A)^{1/2} & A^* \end{bmatrix}. \tag{2.8}$$

(Check that this is a unitary element of $\mathbb{M}_{2n}$.)
Now let Ψ be the compression map taking a $2n \times 2n$ matrix to its top left $n \times n$ corner. Then Ψ is positive and unital. So, the composition $\Phi \circ \Psi$ is positive and unital. Now Choi's inequality (2.6) can be used to get

$$[\,(\Phi \circ \Psi)(\hat{A})\,]\,[\,(\Phi \circ \Psi)(\hat{A}^*)\,] \leq (\Phi \circ \Psi)(I).$$

But this says

$$\Phi(A)\Phi(A^*) \leq I.$$

This shows that $\|\Phi(A)\| \leq 1$ whenever $\|A\| \leq 1$. Hence, $\|\Phi\| = 1$. ■

We can extend the result to any positive linear map as follows.

2.3.8 Corollary

Let Φ be a positive linear map. Then $\|\Phi\| = \|\Phi(I)\|$.

Proof. Let $P = \Phi(I)$, and assume first that P is invertible. Let

$$\Psi(A) = P^{-1/2}\Phi(A)P^{-1/2}.$$

Then Ψ is a positive unital linear map. So, we have

$$\|\Phi(A)\| = \|P^{1/2}\Psi(A)P^{1/2}\| \leq \|P\|\;\|\Psi(A)\| \leq \|P\|\;\|A\|.$$

Thus $\|\Phi\| \le \|P\|$; and since $\Phi(I) = P$, we have $\|\Phi\| = \|P\|$. This proves the assertion when $\Phi(I)$ is invertible. The general case follows from this by considering the family $\Phi_\varepsilon(A) = \Phi(A) + \varepsilon I$ and letting $\varepsilon \downarrow 0$. ■

The assertion of (this Corollary to) the Russo-Dye theorem is some times phrased as: *every positive linear map on $\mathbb{M}_n$ attains its norm at the identity matrix.*

2.3.9 Exercise

There is a simpler proof of this theorem in the case of positive linear functionals. In this case $\varphi(A) = \operatorname{tr} AX$ for some positive matrix X. Then

$$|\varphi(A)| = |\operatorname{tr} AX| \le \|A\| \ \|X\|_1 = \|A\| \ \operatorname{tr} X = \varphi(I) \ \|A\|.$$

Here $\|T\|_1$ is the trace norm of T defined as $\|T\|_1 = s_1(T) + \cdots + s_n(T)$. The inequality above is a consequence of the fact that this norm is the dual of the norm $\|\cdot\|$.

2.4 SOME APPLICATIONS

We have seen several examples of positive maps. Using the Russo-Dye Theorem we can calculate their norms easily. Thus, for example,

$$\|\mathcal{C}(A)\| \ \le \ \|A\| \tag{2.9}$$

for every pinching of A. (This can be proved in several ways. See MA pp. 50, 97.)

If A is positive, then the Schur multiplier S_A is a positive map. So,

$$\|S_A\| \ = \ \|S_A(I)\| \ = \ \|A \circ I\| = \max a_{ii}. \tag{2.10}$$

This too can be proved in many ways. We have seen this before in Theorem 1.4.1.

We have discussed the Lyapunov equation

$$A^*X + XA = W, \tag{2.11}$$

where A is an operator whose spectrum is contained in the open right half plane. (Exercise 1.2.10, Example 2.2.1 (x)). Solving this equation means finding the inverse of the Lyapunov operator L_A defined as $L_A(X) = A^*X + XA$. We have seen that L_A^{-1} is a positive linear map. In some situations W is known with some imprecision, and we have the perturbed equation

$$A^*X + XA = W + \triangle W. \tag{2.12}$$

If X and $X+\triangle X$ are the solutions to (2.11) and (2.12), respectively, one wants to find bounds for $\|\triangle X\|$. This is a very typical problem in numerical analysis. Clearly,

$$\|\triangle X\| \;\leq\; \|L_A^{-1}\| \; \|\triangle W\|.$$

Since L_A^{-1} is positive we have $\|L_A^{-1}\| = \|L_A^{-1}(I)\|$. This simplifies the problem considerably. The same considerations apply to the Stein equation (Exercise 1.2.11).

Let $\otimes^k\mathcal{H}$ be the k-fold tensor product $\mathcal{H}\otimes\cdots\otimes\mathcal{H}$ and let $\otimes^k A$ be the k-fold product $A\otimes\cdots\otimes A$ of an operator A on $\mathcal{H}$. For $1 \leq k \leq n$, let $\wedge^k\mathcal{H}$ be the subspace of $\otimes^k\mathcal{H}$ spanned by antisymmetric tensors. This is called the *antisymmetric tensor product*, *exterior product*, or *Grassmann product*. The operator $\otimes^k A$ leaves this space invariant and the restriction of $\otimes^k A$ to it is denoted as $\wedge^k A$. This is called the kth *Grassmann power*, or the *exterior power* of A.

Consider the map $A \longmapsto \otimes^k A$. The *derivative* of this map at A, denoted as $D\otimes^k(A)$, is a linear map from $\mathcal{L}(\mathcal{H})$ into $\mathcal{L}(\otimes^k\mathcal{H})$. Its action is given as

$$D\otimes^k(A)(B) = \frac{d}{dt}\bigg|_{t=0} \otimes^k (A+tB).$$

Hence,

$$D\otimes^k(A)(B) = B\otimes A\otimes\cdots\otimes A + A\otimes B\otimes\cdots\otimes A + \cdots + A\otimes\cdots\otimes A\otimes B. \tag{2.13}$$

It follows that

$$\|D\otimes^k(A)\| \;=\; k\|A\|^{k-1}. \tag{2.14}$$

We want to find an expression for $\|D\wedge^k(A)\|$.

Recall that $\wedge^k$ is multiplicative, $*$ - preserving, and unital (but not linear!). Let $A = USV$ be the singular value decomposition of A. Then

$$\begin{aligned} D\wedge^k(A)(B) &= \left.\frac{d}{dt}\right|_{t=0} \wedge^k (A+tB) \\ &= \left.\frac{d}{dt}\right|_{t=0} \wedge^k (USV+tB) \\ &= \left.\frac{d}{dt}\right|_{t=0} \wedge^k \; U\cdot\wedge^k(S+tU^*BV^*)\cdot\wedge^k V \\ &= \wedge^k U\left[\left.\frac{d}{dt}\right|_{t=0} \wedge^k (S+tU^*BV^*)\right]\wedge^k V. \end{aligned}$$

Thus

$$\|D\wedge^k(A)(B)\| = \|D\wedge^k(S)(U^*BV^*)\|,$$

and hence

$$\|D\wedge^k(A)\| = \|D\wedge^k(S)\|.$$

Thus to calculate $\|D\wedge^k(A)\|$, we may assume that A is positive and diagonal.

Now note that if A is positive, then for every positive B, the expression (2.13) is positive. So $D\otimes^k(A)$ is a positive linear map from $\mathcal{L}(\mathcal{H})$ into $\mathcal{L}(\otimes^k\mathcal{H})$. The operator $D\wedge^k(A)(B)$ is the restriction of (2.13) to the invariant subspace $\wedge^k\mathcal{H}$. So $\wedge^k(A)$ is also a positive linear map. Hence

$$\|D\wedge^k(A)\| = \|D\wedge^k(A)(I)\|.$$

Let $A = \operatorname{diag}(s_1,\dots,s_n)$ with $s_1 \ge s_2 \ge \cdots \ge s_n \ge 0$. Then $\wedge^k A$ is a diagonal matrix of size $\binom{n}{k}$ whose diagonal entries are $s_{i_1}s_{i_2}\cdots s_{i_k}$, $1 \le i_1 < i_2 < \cdots < i_k \le n$. Use this to see that

$$\|D\wedge^k(A)\| = p_{k-1}(s_1,\dots,s_k) \tag{2.15}$$

the elementary symmetric polynomial of degree $k-1$ with arguments $s_1, \ldots, s_k$.

The effect of replacing $D \wedge^k (A)(B)$ by $D \wedge^k (A)(I)$ is to reduce a highly noncommutative problem to a simple commutative one. Another example of this situation is given in Section 2.7.

2.5 THREE QUESTIONS

Let $\Phi : \mathbb{M}_n \longmapsto \mathbb{M}_k$ be a linear map. We have seen that if Φ is positive, then

$$\|\Phi\| = \|\Phi(I)\|. \tag{2.16}$$

Clearly, this is a useful and important theorem. It is natural to explore how much, and in what directions, it can be extended.

Question 1 Are there linear maps other than positive ones for which (2.16) is true? In other words, if a linear map Φ attains its norm at the identity, then must Φ be positive?

Before attempting an answer, we should get a small irritant out of the way. If the condition (2.16) is met by Φ, then it is met by $-\Phi$ also. Clearly, both of them cannot be positive maps. So assume Φ satisfies (2.16) and

$$\Phi(I) \geq O. \tag{2.17}$$

2.5.1 Exercise

If $k = 1$, the answer to our question is yes. In this case $\varphi(A) = \text{tr}AX$ for some X. Then $\|\varphi\| = \|X\|_1$ (see Exercise 2.3.9). So, if φ satisfies (2.16) and (2.17), then $\|X\|_1 = \text{tr}X$. Show that this is true if and only if X is positive. Hence φ is positive.

If $k \geq 2$, this is no longer true. For example, let Φ be the linear map on $\mathbb{M}_2$ defined as

$$\Phi\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) = \begin{bmatrix} a_{11} & a_{12} \\ 0 & 0 \end{bmatrix}.$$

Then $\|\Phi\| = \|\Phi(I)\| = 1$ and $\Phi(I) \geq O$, but Φ is not positive.

It is a remarkable fact that if Φ is unital and $\|\Phi\| = 1$, then Φ is positive. Thus supplementing (2.16) with the condition $\Phi(I) = I$ ensures that Φ is positive. This is proved in the next section.

Question 2 Let $\mathcal{S}$ be a linear subspace of $\mathbb{M}_n$ and let $\Phi : \mathcal{S} \to \mathbb{M}_k$ be a linear map. Do we still have a theorem like the Russo-Dye theorem? In other words how crucial is the fact that the domain of Φ is $\mathbb{M}_n$ (or possibly a subalgebra)?

Again, for the question to be meaningful, we have to demand of $\mathcal{S}$ a little more structure. If we want to talk of positive unital maps, then $\mathcal{S}$ must contain some positive elements and I.

2.5.2 Definition

A linear subspace $\mathcal{S}$ of $\mathbb{M}_n$ is called an *operator system* if it is $*$ closed (i.e., if $A \in \mathcal{S}$, then $A^* \in \mathcal{S}$) and contains I.

Let $\mathcal{S}$ be an operator system. We want to know whether a positive linear map $\Phi : \mathcal{S} \longmapsto \mathbb{M}_k$ attains its norm at I. The answer is yes if $k = 1$, and no if $k \geq 2$. However, we do have $\|\Phi\| \leq 2\|\Phi(I)\|$ for all k.

A related question is the following:

Question 3 By the Hahn-Banach theorem, every linear functional φ on (a linear subspace) $\mathcal{S}$ can be extended to a linear functional $\widehat{\varphi}$ on $\mathbb{M}_n$ in such a way that $\|\widehat{\varphi}\| = \|\varphi\|$. Now we are considering positivity rather than norms. So we may ask whether a positive linear functional φ on an operator system $\mathcal{S}$ in $\mathbb{M}_n$ can be extended to a positive linear functional $\widehat{\varphi}$ on $\mathbb{M}_n$. The answer is yes. This is called the *Krein extension theorem.* Then since $\|\widehat{\varphi}\| = \varphi(I)$, we have $\|\varphi\| = \varphi(I)$.

Next we may ask whether a positive linear map Φ from $\mathcal{S}$ into $\mathbb{M}_k$ can be extended to a positive linear map $\widehat{\Phi}$ from $\mathbb{M}_n$ into $\mathbb{M}_k$. If this were the case, then we would have $\|\Phi\| = \|\Phi(I)\|$. But we have said that this is not always true when $k \geq 2$. This is illustrated by the following example.

2.5.3 Example

Let n be any number bigger than 2 and let S be the $n \times n$ permutation matrix

$$S = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \end{bmatrix}.$$

Let $\mathcal{S}$ be the collection of all matrices C of the form $C = aI + bS + cS^*$, $a, b, c \in \mathbb{C}$. (The matrices C are *circulant* matrices.) Then $\mathcal{S}$ is an operator system in $\mathbb{M}_n$. What are the positive elements of $\mathcal{S}$? First, we must have $a \geq 0$ and $c = \overline{b}$. The eigenvalues of S are $1, \omega, \dots, \omega^{n-1}$, the n roots of 1. So, the eigenvalues of C are

$$a + b + \overline{b},\ a + b\omega + \overline{b\omega},\ \dots, a + b\omega^{n-1} + \overline{b\omega^{n-1}},$$

and C is positive if and only if all these numbers are nonnegative.

It is helpful to consider the special case $n = 4$. The fourth roots of unity are $\{1, i, -1, -i\}$ and one can see that a matrix C of the type above is positive if and only if

$$a \geq 2\,|\text{Re}\ b| \quad \text{and} \quad a \geq 2\,|\text{Im}\ b|.$$

Let $\Phi : \mathcal{S} \to \mathbb{M}_2$ be the map defined as

$$\Phi(C) = \begin{bmatrix} a & \sqrt{2}b \\ \sqrt{2}c & a \end{bmatrix}.$$

Then Φ is linear, positive, and unital. Since

$$\Phi(S) = \begin{bmatrix} 0 & \sqrt{2} \\ 0 & 0 \end{bmatrix},$$

$\|\Phi\| \geq \sqrt{2}$. So, Φ cannot be extended to a positive, linear, unital map from $\mathbb{M}_4$ into $\mathbb{M}_2$.

2.5.4 Exercise

Let $n \geq 3$ and consider the operator system $\mathcal{S} \subset \mathbb{M}_n$ defined in the example above. For every t the map $\Phi : \mathcal{S} \to \mathbb{M}_2$ defined as

$$\Phi(C) = \begin{bmatrix} a & tb \\ tc & a \end{bmatrix}$$

is linear and unital. Show that for $1 < t < 2$ there exists an n such that the map Φ is positive.

We should remark here that the elements of $\mathcal{S}$ commute with each other (though, of course, $\mathcal{S}$ is not a subalgebra of $\mathbb{M}_n$).

In the next section we prove the statements that we have made in answer to the three questions.

2.6 POSITIVE MAPS ON OPERATOR SYSTEMS

Let $\mathcal{S}$ be an operator system in $\mathbb{M}_n$, $\mathcal{S}_{s.a.}$ the set of self-adjoint elements of $\mathcal{S}$, and $\mathcal{S}_+$ the set of positive elements in it.

Some of the operations that we performed freely in $\mathbb{M}_n$ may take us outside $\mathcal{S}$. Thus if $T \in \mathcal{S}$, then $\operatorname{Re} T = \frac{1}{2}(T + T^*)$ and $\operatorname{Im} T = \frac{1}{2i}(T - T^*)$ are in $\mathcal{S}$. However, if $A \in \mathcal{S}_{s.a.}$, then the positive and negative parts $A_\pm$ in the Jordan decomposition of A need not be in $\mathcal{S}_+$. For example, consider

$$\mathcal{S} = \{A \in \mathbb{M}_3 : a_{11} = a_{22} = a_{33}\}.$$

This is an operator system. The matrix $A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ is in $\mathcal{S}$. Its Jordan components are

$$A_+ = \frac{1}{2}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \quad A_- = \frac{1}{2}\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}.$$

They do not belong to $\mathcal{S}$.

However, it is possible still to write every Hermitian element A of $\mathcal{S}$ as

$$A = P_+ - P_- \quad \text{where } P_\pm \in \mathcal{S}_+. \tag{2.18}$$

Just choose

$$P_\pm = \frac{\|A\|I \pm A}{2}. \tag{2.19}$$

Thus we can write every $T \in \mathcal{S}$ as

$$T = A + iB \qquad\qquad (A, B \in \mathcal{S}_{s.a})$$

$$= (P_+ - P_-) + i(Q_+ - Q_-) \qquad (P_\pm, Q_\pm \in \mathcal{S}_+).$$

Using this decomposition we can prove the following lemma.

2.6.1 Lemma

Let Φ be a positive linear map from an operator system $\mathcal{S}$ into $\mathbb{M}_k$. Then $\Phi(T^*) = \Phi(T)^*$ for all $T \in \mathcal{S}$.

2.6.2 Exercise

If $A = P_1 - P_2$ where P_1, P_2 are positive, then

$$\|A\| \leq \max(\|P_1\|, \|P_2\|).$$

2.6.3 Theorem

Let Φ be a positive linear map from an operator system $\mathcal{S}$ into $\mathbb{M}_k$. Then

(i) $\|\Phi(A)\| \leq \|\Phi(I)\|\|A\|$ for all $A \in \mathcal{S}_{s.a.}$

and

(ii) $\|\Phi(T)\| \leq 2\|\Phi(I)\|\|T\|$ for all $T \in \mathcal{S}$.

(Thus if Φ is also unital, then $\|\Phi\| = 1$ on the space $\mathcal{S}_{s.a.}$, and $\|\Phi\| \leq 2$ on $\mathcal{S}$.)

Proof. If P is a positive element of $\mathcal{S}$, then $O \leq P \leq \|P\|I$, and hence $O \leq \Phi(P) \leq \|P\|\Phi(I)$.

If A is a Hermitian element of $\mathcal{S}$, use the decomposition (2.18), Exercise 2.6.2, and the observation of the preceding sentence to see that

$$\begin{aligned}\|\Phi(A)\| &= \|\Phi(P_+) - \Phi(P_-)\| \\ &\leq \max(\|\Phi(P_+)\|, \|\Phi(P_-)\|) \\ &\leq \max(\|P_+\|, \|P_-\|)\, \|\Phi(I)\| \\ &\leq \|A\|\ \|\Phi(I)\|.\end{aligned}$$

This proves the first inequality of the theorem. The second is obtained from this by using the Cartesian decomposition of T. ■

Exercise 2.5.4 shows that the factor 2 in the inequality (ii) of Theorem 2.6.3 is unavoidable in general. It can be dropped when $k = 1$:

2.6.4 Theorem

Let φ be a positive linear functional on an operator system $\mathcal{S}$. Then $\|\varphi\| = \varphi(I)$.

Proof. Let $T \in \mathcal{S}$ and $\|T\| \leq 1$. We want to show $|\varphi(T)| \leq \varphi(I)$. If $\varphi(T)$ is the complex number $re^{i\theta}$, we may multiply T by $e^{-i\theta}$, and thus assume $\varphi(T)$ is real and positive. So, if $T = A + iB$ in the Cartesian decomposition, then $\varphi(T) = \varphi(A)$. Hence by part (i) of Theorem 2.6.3 $\varphi(T) \leq \varphi(I)\|A\| \leq \varphi(I)\|T\|$. ■

The converse is also true.

2.6.5 Theorem

Let φ be a linear functional on $\mathcal{S}$ such that $\|\varphi\| = \varphi(I)$. Then φ is positive.

Proof. Assume, without loss of generality, that $\varphi(I) = 1$. Let A be a positive element of $\mathcal{S}$ and let $\sigma(A)$ be its spectrum. Let $a = \min \sigma(A)$ and $b = \max \sigma(A)$. We will show that the point $\varphi(A)$ lies in the interval $[a, b]$. If this is not the case, then there exists a disk $D = \{z : |z - z_0| \leq r\}$ such that $\varphi(A)$ is outside D but D contains $[a, b]$, and hence also $\sigma(A)$. From the latter condition it follows that $\sigma(A - z_0 I)$ is contained in the disk $\{z : |z| \leq r\}$, and hence $\|A - z_0 I\| \leq r$. Using the conditions $\|\varphi\| = \varphi(I) = 1$, we get from this

$$|\varphi(A) - z_0| = |\varphi(A - z_0\, I)| \leq \|\varphi\|\ \|A - z_0\, I\| \leq r.$$

But then $\varphi(A)$ could not have been outside D.

This shows that $\varphi(A)$ is a nonnegative number, and the theorem is proved. ■

2.6.6 Theorem (The Krein Extension Theorem)

Let $\mathcal{S}$ be an operator system in $\mathbb{M}_n$. Then every positive linear functional on $\mathcal{S}$ can be extended to a positive linear functional on $\mathbb{M}_n$.

Proof. Let φ be a positive linear functional on $\mathcal{S}$. By Theorem 2.6.4, $\|\varphi\| = \varphi(I)$. By the Hahn-Banach Theorem, φ can be extended to a linear functional $\widehat{\varphi}$ on $\mathbb{M}_n$ with $\|\widehat{\varphi}\| = \|\varphi\| = \varphi(I)$. But then $\widehat{\varphi}$ is positive by Theorem 2.6.5 (or by Exercise 2.5.1). ■

Finally we have the following theorem that proves the assertion made at the end of the discussion of Question 1 in Section 2.5.

2.6.7 Theorem

Let $\mathcal{S}$ be an operator system and let $\Phi : \mathcal{S} \longrightarrow \mathbb{M}_k$ be a unital linear map such that $\|\Phi\| = 1$. Then Φ is positive.

Proof. For each unit vector x in $\mathbb{C}^k$, let

$$\varphi_x(A) = \langle x, \Phi(A)x \rangle, \;\; A \in \mathcal{S}.$$

This is a unital linear functional on $\mathcal{S}$. Since $|\varphi_x(A)| \leq \|\Phi(A)\| \leq \|A\|$, we have $\|\varphi_x\| = 1$. So, by Theorem 2.6.5, φ_x is positive. In other words, if A is positive, then for every unit vector x

$$\varphi_x(A) \;=\; \langle x, \Phi(A)x \rangle \;\geq 0.$$

But that means Φ is positive. ■

2.7 SUPPLEMENTARY RESULTS AND EXERCISES

Some of the theorems in Section 2.3 are extended in various directions in the following propositions.

2.7.1 Proposition

Let Φ be a positive unital linear map on $\mathbb{M}_n$ and let A be a positive matrix. Then

$$\Phi(A)^r \geq \Phi(A^r) \quad \text{for } 0 \leq r \leq 1.$$

Proof. Let $0 < r < 1$. Using the integral representation (1.39) we have

$$A^r = \int_0^\infty A(\lambda + A)^{-1} d\mu(\lambda),$$

where μ is a positive measure on $(0, \infty)$. So it suffices to show that

$$\Phi(A)(\lambda + \Phi(A))^{-1} \geq \Phi(A(\lambda + A)^{-1})$$

for all $\lambda > 0$. We have the identity

$$A(\lambda + A)^{-1} = I - \lambda(\lambda + A)^{-1}.$$

Apply Φ to both sides and use Theorem 2.3.6 to get

$$\begin{aligned}\Phi(A(\lambda + A)^{-1}) &= I - \lambda\Phi((\lambda + A)^{-1}) \\ &\leq I - \lambda(\Phi(\lambda + A))^{-1} \\ &= I - \lambda(\lambda + \Phi(A))^{-1}.\end{aligned}$$

The identity stated above shows that the last expression is equal to $\Phi(A)(\lambda + \Phi(A))^{-1}$. ■

2.7.2 Exercise

Let Φ be a positive unital linear map on $\mathbb{M}_n$ and let A be a positive matrix. Show that

$$\Phi(A)^r \leq \Phi(A^r)$$

if $1 \leq r \leq 2$. If A is strictly positive, then this is true also when $-1 \leq r \leq 0$. [Hint: Use integral representations of A^r as in Theorem 1.5.8, Exercise 1.5.10, and the inequalities (2.5) and (2.7).]

2.7.3 Proposition

Let Φ be a strictly positive linear map on $\mathbb{M}_n$. Then

$$\Phi(HA^{-1}H) \;\geq\; \Phi(H)\;\Phi(A)^{-1}\;\Phi(H) \tag{2.20}$$

whenever H is Hermitian and $A > 0$.

Proof. Let

$$\Psi(Y) \;=\; \Phi(A)^{-1/2}\;\Phi(A^{1/2}YA^{1/2})\;\Phi(A)^{-1/2}. \tag{2.21}$$

Then Ψ is positive and unital. By Kadison's inequality we have $\Psi(Y^2) \geq \Psi(Y)^2$ for every Hermitian Y. Choose $Y = A^{-1/2}HA^{-1/2}$ to get

$$\Psi(A^{-1/2}HA^{-1}HA^{-1/2}) \;\geq\; \Big(\Psi(A^{-1/2}HA^{-1/2})\Big)^2.$$

Use (2.21) now to get (2.20). ■

2.7.4 Exercise

Construct an example to show that a more general version of (2.20)

$$\Phi(X^* A^{-1} X) \;\geq\; \Phi(X)^* \Phi(A)^{-1} \Phi(X),$$

where X is arbitrary and A positive, is not always true.

2.7.5 Proposition

Let Φ be a strictly positive linear map on $\mathbb{M}_n$ and let $A > O$. Then

$$A \geq X^* A^{-1} X \implies \Phi(A) \geq \Phi(X)^* \Phi(A)^{-1} \Phi(X). \tag{2.22}$$

Proof. Let Ψ be the linear map defined by (2.21). By the Russo-Dye theorem

$$Y^* Y \leq I \implies \Psi(Y)^* \Psi(Y) \leq I.$$

Let $A \geq X^* A^{-1} X$ and put $Y = A^{-1/2} X A^{-1/2}$. Then $Y^* Y = A^{-1/2} X^* A^{-1} X A^{-1/2} \leq I$. Hence $\Psi(A^{-1/2} X^* A^{-1/2}) \Psi(A^{-1/2} X A^{-1/2}) \leq I$. Use (2.21) again to get (2.22). ■

In classical probability the quantity

$$\mathrm{var}(f) \;=\; Ef^2 - (Ef)^2 \tag{2.23}$$

is called the *variance* of the real function f. In analogy we consider

$$\mathrm{var}(A) \;=\; \Phi(A^2) - \Phi(A)^2_{;} \tag{2.24}$$

where A is Hermitian and Φ a positive unital linear map on $\mathbb{M}_n$. Kadison's inequality says $\mathrm{var}(A) \geq O$. The following proposition gives an upper bound for $\mathrm{var}(A)$.

2.7.6 Proposition

Let Φ be a positive unital linear map and let A be a Hermitian operator with $mI \leq A \leq MI$. Then

$$\mathrm{var}(A) \;\leq\; (MI - \Phi(A))(\Phi(A) - mI) \;\leq\; \frac{1}{4}(M-m)^2 I. \tag{2.25}$$

Proof. The matrices $MI-A$ and $A-mI$ are positive and commute with each other. So, $(MI-A)(A-mI) \geq O$; i.e.,

$$A^2 \;\leq\; MA + mA - MmI.$$

Apply Φ to both sides and then subtract $\Phi(A)^2$ from both sides. This gives the first inequality in (2.25). To prove the second inequality note that if $m \leq x \leq M$, then $(M-x)(x-m) \leq \frac{1}{4}(M-m)^2$. ∎

2.7.7 Exercise

Let $x \in \mathbb{C}^n$. We say $x \geq 0$ if all its coordinates x_j are nonnegative. Let $e = (1, \ldots, 1)$.

A matrix S is called *stochastic* if $s_{ij} \geq 0$ for all i, j, and $\sum\limits_{j=1}^{n} s_{ij} = 1$ for all i. Show that S is stochastic if and only if

$$x \geq 0 \implies Sx \geq 0 \tag{2.26}$$

and

$$Se \;=\; e. \tag{2.27}$$

The property (2.26) can be described by saying that the linear map defined by S on $\mathbb{C}^n$ is *positive*, and (2.27) by saying that S is *unital.*

If x is a real vector, let $x^2 = (x_1^2, \ldots, x_n^2)$. Show that if S is a stochastic matrix and $m \leq x_j \leq M$, then

$$0 \leq S(x^2) - S(x)^2 \leq (Me - Sx)(Sx - me) \leq \frac{1}{4}(M-m)^2 e. \tag{2.28}$$

A special case of this is obtained by choosing $s_{ij} = \frac{1}{n}$ for all i, j. If $\overline{x} = \frac{1}{n}\sum x_j$, this gives

$$\frac{1}{n}\sum_{j=1}^{n}(x_j - \overline{x})^2 \;\leq\; (M - \overline{x})(\overline{x} - m) \;\leq\; \frac{1}{4}(M-m)^2\,. \tag{2.29}$$

An inequality complementary to (2.7) is given by the following proposition.

2.7.8 Proposition

Let Φ be strictly positive and unital. Let $0 < m < M$. Then for every strictly positive matrix A with $mI \le A \le MI$, we have

$$\Phi(A^{-1}) \;\le\; \frac{(M+m)^2}{4Mm}\,\Phi(A)^{-1}\,. \tag{2.30}$$

Proof. The matrices $A - mI$ and $MA^{-1} - I$ are positive and commute with each other. So, $O \le (A - mI)(MA^{-1} - I)$. This gives

$$MmA^{-1} \;\le\; (M+m)I - A,$$

and hence

$$Mm\Phi(A^{-1}) \;\le\; (M+m)I - \Phi(A).$$

Now, if c and x are real numbers, then $(c - 2x)^2 \ge 0$ and therefore, for positive x we have $c - x \le \frac{1}{4}c^2x^{-1}$. So, we get

$$Mm\Phi(A^{-1}) \;\le\; \frac{(M+m)^2}{4}\Phi(A)^{-1}. \;\blacksquare$$

A very special corollary of this is the inequality

$$\langle x, Ax\rangle\,\langle x, A^{-1}x\rangle \;\le\; \frac{(M+m)^2}{4Mm}, \tag{2.31}$$

for every unit vector x. This is called the *Kantorovich inequality.*

2.7.9 Exercise

Let f be a convex function on an interval $[m, M]$ and let L be the linear interpolant

$$L(t) \;=\; \frac{1}{M-m}\left[(t-m)f(M) + (M-t)f(m)\right].$$

Show that if Φ is a unital positive linear map, then for every Hermitian matrix A whose spectrum is contained in $[m, M]$, we have

$$\Phi(f(A)) \ \leq \ L(\Phi(A)).$$

Use this to obtain Propositions 2.7.6 and 2.7.8.

The space $\mathbb{M}_n$ has a natural inner product defined as

$$\langle A, B \rangle \ = \ \text{tr } A^*B. \tag{2.32}$$

If Φ is a linear map on $\mathbb{M}_n$, we define its *adjoint* Φ^* as the linear map that satisfies the condition

$$\langle \Phi(A), B \rangle \ = \ \langle A, \Phi^*(B) \rangle \text{ for all } A, B. \tag{2.33}$$

2.7.10 Exercise

The linear map Φ is positive if and only if Φ^* is positive. Φ is unital if and only if Φ^* is trace preserving; i.e., tr $\Phi^*(A) = \text{tr } A$ for all A.

We say Φ is a *doubly stochastic* map on $\mathbb{M}_n$ if it is positive,unital, and trace preserving (i.e., both Φ and Φ^* are positive and unital).

2.7.11 Exercise

(i) Let Φ be the linear map on $\mathbb{M}_n$ defined as $\Phi(A) = X^*AX$. Show that $\Phi^*(A) = XAX^*$.

(ii) For any A, let $S_A(X) = A \circ X$ be the Schur product map. Show that $(S_A)^* = S_{A^*}$.

(iii) Every pinching is a doubly stochastic map.

(iv) Let $L_A(X) = A^*X + XA$ be the Lyapunov operator, where A is a matrix with its spectrum in the open right half plane. Show that $(L_A^{-1})^* = (L_{A^*})^{-1}$.

A norm $|||\cdot|||$ on $\mathbb{M}_n$ is said to be *unitarily invariant* if $|||UAV||| = |||A|||$ for all A and unitary U, V. It is convenient to make a normalisation so that $|||A||| = 1$ whenever A is a rank-one orthogonal projection.

Special examples of such norms are the *Ky Fan norms*

$$||A||_{(k)} = \sum_{j=1}^{k} s_j(A), \ 1 \le k \le n,$$

and the *Schatten p-norms*

$$||A||_p = \left[\sum_{j=1}^{n} (s_j(A))^p \right]^{1/p}, \ 1 \le p \le \infty.$$

Note that the operator norm, in this notation, is

$$||A|| = ||A||_\infty = ||A||_{(1)},$$

and the trace norm is the norm

$$||A||_1 = ||A||_{(n)}.$$

The norm $\|A\|_2$ is also called the *Hilbert-Schmidt norm.*

The following facts are well known:

$$||A||_{(k)} = \min\{||B||_{(n)} + k||C|| \ : \ A = B + C\}. \tag{2.34}$$

If $||A||_{(k)} \le ||B||_{(k)}$ for $1 \le k \le n$, then $|||A||| \le |||B|||$ for all unitarily invariant norms. This is called the *Fan dominance theorem.* (See MA, p. 93.)

For any three matrices A, B, C we have

$$|||ABC||| \le ||A|| \ |||B||| \ ||C||. \tag{2.35}$$

If Φ is a linear map on $\mathbb{M}_n$ and $|||\cdot|||$ any unitarily invariant norm, then we use the notation $|||\Phi|||$ for

$$|||\Phi||| = \sup_{|||A|||=1} |||\Phi(A)||| = \sup_{|||A|||\le 1} |||\Phi(A)|||. \tag{2.36}$$

In the same way,

$$||\Phi||_1 = \sup_{||A||_1=1} ||\Phi(A)||_1,$$

etc.
The norm $||A||_1$ is the dual of the norm $||A||$ on M_n. Hence

$$||\Phi|| \;=\; ||\Phi^*||_1. \tag{2.37}$$

2.7.12 Exercise

Let $|||\cdot|||$ be any unitarily invariant norm on $\mathbb{M}_n$.

(i) Use the relations (2.34) and the Fan dominance theorem to show that if $\|\Phi\| \leq 1$ and $\|\Phi^*\| \leq 1$, then $|||\Phi||| \leq 1$.

(ii) If Φ is a doubly stochastic map, then $|||\Phi||| \leq 1$.

(iii) If $A \geq O$, then $|||A \circ X||| \leq \max\ a_{ii}|||X|||$ for all X.

(iv) Let L_A be the Lyapunov operator associated with a positively stable matrix A. We know that $||L_A^{-1}|| = ||L_A^{-1}(I)||$. Show that in the special case when A is normal we have $|||L_A^{-1}||| = ||L_A^{-1}(I)||$ $= [2 \min \operatorname{Re} \lambda_i]^{-1}$, where λ_i are the eigenvalues of A.

2.7.13 Exercise

Let A and B be Hermitian matrices. Suppose $A = \Phi(B)$ for some doubly stochastic map Φ on $\mathbb{M}_n$. Show that A is a convex combination of unitary conjugates of B; i.e., there exist unitary matrices $U_1, \ldots, U_k$ and positive numbers $p_1, \ldots, p_k$ with $\sum p_j = 1$ such that

$$A = \sum_{j=1}^{k} p_j U_j^* B U_j.$$

[Hints: There exist diagonal matrices D_1 and D_2, and unitary matrices W and V such that $A = W^* D_1 W$ and $B = V D_2 V^*$. Use this to show that $D_1 = \Psi(D_2)$ where Ψ is a doubly stochastic map. By Birkhoff's theorem there exist permutation matrices $S_1, \ldots, S_k$ and positive numbers $p_1, \ldots, p_k$ with $\sum p_j = 1$ such that

$$D_1 = \sum_{j=1}^{k} p_j S_j^* D_2 S_j.$$

Choose $U_j = V S_j W$. (Note that the matrices U_j and the numbers p_j depend on Φ, A and B.)]

Let $\mathbb{H}_n$ be the set of all $n \times n$ Hermitian matrices. This is a real vector space. Let I be an open interval and let $\mathrm{C}^1(I)$ be the space of continuously differentiable real functions on I. Let $\mathbb{H}_n(I)$ be the set of all Hermitian matrices whose eigenvalues belong to I. This is an open subset of $\mathbb{H}_n$. Every function f in $\mathrm{C}^1(I)$ induces a map $A \longmapsto f(A)$ from $\mathbb{H}_n(I)$ into $\mathbb{H}_n$. This induced map is differentiable and its derivative is given by an interesting formula known as the *Daleckii-Krein formula.*

For each $A \in \mathbb{H}_n(I)$ the derivative $Df(A)$ at A is a linear map from $\mathbb{H}_n$ into itself. If $A = \sum \lambda_i P_i$ is the spectral decomposition of A, then the formula is

$$Df(A)(B) = \sum_i \sum_j \frac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j} P_i B P_j \tag{2.38}$$

for every $B \in \mathbb{H}_n$. For $i = j$, the quotient in (2.38) is to be interpreted as $f'(\lambda_i)$.

This formula can be expressed in another way. Let $f^{[1]}$ be the function on $I \times I$ defined as

$$\begin{aligned} f^{[1]}(\lambda, \mu) &= \frac{f(\lambda) - f(\mu)}{\lambda - \mu} \text{ if } \lambda \neq \mu, \\ f^{[1]}(\lambda, \lambda) &= f'(\lambda). \end{aligned}$$

This is called the *first divided difference* of f. For $A \in \mathbb{H}_n(I)$, let $f^{[1]}(A)$ be the $n \times n$ matrix

$$f^{[1]}(A) = \left[\left[\frac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j}\right]\right], \tag{2.39}$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A. The formula (2.38) says

$$Df(A)(B) = f^{[1]}(A) \circ B, \tag{2.40}$$

where $\circ$ denotes the Schur product taken in a basis in which A is diagonal. A proof of this is given in Section 5.3.

Suppose a real function f on an interval I has the following property: if A and B are two elements of $\mathbb{H}_n(I)$ and $A \geq B$, then $f(A) \geq f(B)$. We say that such a function f is *matrix monotone of order n*

on I. If f is matrix monotone of order n for all $n = 1, 2, \dots$, then we say f is *operator monotone.*

Matrix convexity of order n and operator convexity can be defined in a similar way. In Chapter 1 we have seen that the function $f(t) = t^2$ on the interval $[0, \infty)$ is not matrix monotone of order 2, and the function $f(t) = t^3$ is not matrix convex of order 2. We have seen also that the function $f(t) = t^r$ on the interval $[0, \infty)$ is operator monotone for $0 \le r \le 1$, and it is operator convex for $1 \le r \le 2$ and for $-1 \le r \le 0$. More properties of operator monotone and convex functions are studied in Chapters 4 and 5.

It is not difficult to prove the following, using the formula (2.40).

2.7.14 Exercise

If a function $f \in \mathrm{C}^1(I)$ is matrix monotone of order n, then for each $A \in \mathbb{H}_n(I)$, the matrix $f^{[1]}(A)$ defined in (2.39) is positive.

The converse of this statement is also true. A proof of this is given in Section 5.3. At the moment we note the following interesting consequence of the positivity of $f^{[1]}(A)$.

2.7.15 Exercise

Let $f \in \mathrm{C}^1(I)$ and let f' be the derivative of f. Show that if f is matrix monotone of order n, then for each $A \in \mathbb{H}_n(I)$

$$\|Df(A)\| = \|f'(A)\|. \tag{2.41}$$

By definition

$$\|Df(A)\| = \sup_{\|B\|=1} \|Df(A)(B)\|, \tag{2.42}$$

and

$$Df(A)(B) = \frac{d}{dt}\bigg|_{t=0} f(A + tB).$$

This expression is difficult to calculate for functions such as $f(t) = t^r, 0 < r < 1$. The formula (2.41) gives an easy way to calculate its norm. Its effect is to reduce the supremum in (2.42) to the class of matrices B that commute with A.

2.8 NOTES AND REFERENCES

Since positivity is a useful and interesting property, it is natural to ask what linear transformations preserve it. The variety of interesting examples, and their interpretation as "expectation," make positive linear maps especially interesting. Their characterization, however, has turned out to be slippery, and for various reasons the special class of *completely positive* linear maps has gained in importance.

Among the early major works on positive linear maps is the paper by E. Størmer, *Positive linear maps of operator algebras,* Acta Math., 110 (1963) 233–278. Research expository articles that explain several subtleties include E. Størmer, *Positive linear maps of C^*-algebras,* in *Foundations of Quantum Mechanics and Ordered Linear Spaces,* Lecture Notes in Physics, Vol. 29, Springer, 1974, pp.85–106, and M.-D. Choi, *Positive linear maps,* in *Operator Algebras and Applications, Part 2,* R. Kadison ed., American Math. Soc., 1982. Closer to our concerns are Chapter 2 of V. Paulsen, *Completely Bounded Maps and Operator Algebras,* Cambridge University Press, 2002, and sections of the two reports by T. Ando, *Topics on Operator Inequalities,* Sapporo, 1978 and *Operator-Theoretic Methods for Matrix Inequalities,* Sapporo, 1998.

The inequality (2.5) was proved in the paper R. Kadison, *A generalized Schwarz inequality and algebraic invariants for operator algebras,* Ann. Math., 56 (1952) 494–503. This was generalized by C. Davis, *A Schwarz inequality for convex operator functions,* Proc. Am. Math. Soc., 8 (1957) 42–44, and by M.-D. Choi, *A Schwarz inequality for positive linear maps on C^*-algebras,* Illinois J. Math., 18 (1974) 565–574. The generalizations say that if Φ is a positive unital linear map and f is an operator convex function, then we have a Jensen-type inequality

$$f\Big(\Phi(A)\Big) \leq \Phi\Big(f(A)\Big). \tag{2.43}$$

The inequality (2.7) and the result of Exercise 2.7.2 are special cases of this. Using the integral representation of an operator convex function given in Problem V.5.5 of MA, one can prove the general inequality by the same argument as used in Exercise 2.7.2. The inequality (2.43) characterises operator convex functions, as was noted by C. Davis, *Notions generalizing convexity for functions defined on spaces of matrices,* in *Proc. Symposia Pure Math., Vol. VII, Convexity,* American Math. Soc., 1963.

In our proof of Theorem 2.3.7 we used the fact that any contraction is an average of two unitaries. The infinite-dimensional analogue says that the unit ball of a C^*-algebra is the closed convex hull of the unitary elements. (Unitaries, however, do not constitute the full set of extreme points of the unit ball. See P. R. Halmos, *A Hilbert Space Problem Book,* Second Edition, Springer, 1982.) This theorem about the closed convex hull is also called the *Russo-Dye theorem* and was proved in B. Russo and H. A. Dye, *A note on unitary operators in C^*-algebras,* Duke Math. J., 33 (1966) 413–416.

Applications given in Section 2.4 make effective use of Theorem 2.3.7 in calculating norms of complicated operators. Our discussion of the Lyapunov equation follows the one in R. Bhatia and L. Elsner, *Positive linear maps and the Lyapunov equation,* Oper. Theory: Adv. Appl., 130 (2001) 107–120. As pointed out in this paper, the use of positivity leads to much more economical proofs than those found earlier by engineers. The equality (2.15) was first proved by R. Bhatia and S. Friedland, *Variation of Grassman powers and spectra,* Linear Algebra Appl., 40 (1981) 1–18. The alternate proof using positivity is due to V. S. Sunder, *A noncommutative analogue of* $|DX^k| = |kX^{k-1}|$, ibid., 44 (1982) 87-95. The analogue of the formula (2.15) when the antisymmetric tensor product is replaced by the symmetric one was worked out in R. Bhatia, *Variation of symmetric tensor powers and permanents,* ibid., 62 (1984) 269–276. The harder problem embracing all symmetry classes of tensors was solved in R. Bhatia and J. A. Dias da Silva, *Variation of induced linear operators,* ibid., 341 (2002) 391–402.

Because of our interest in certain kinds of matrix problems involving calculation or estimation of norms we have based our discussion in Section 2.5 on the relation (2.16). There are far more compelling reasons to introduce operator systems. There is a rapidly developing and increasingly important theory of operator spaces (closed linear subspaces of C^*-algebras) and operator systems. See the book by V. Paulsen cited earlier, E. G. Effros and Z.-J. Ruan, *Operator Spaces,* Oxford University Press, 2000, and G. Pisier, *Introduction to Operator Space Theory,* Cambridge University Press, 2003. This is being called the noncommutative or quantized version of Banach space theory. One of the corollaries of the Hahn-Banach theorem is that every separable Banach space is isometrically isomorphic to a subspace of l_∞; and every Banach space is isometrically isomorphic to a subspace of $l_\infty(X)$ for some set X. In the quantized version the commutative space l_∞ is replaced by the noncommutative space $\mathcal{L}(\mathcal{H})$ where $\mathcal{H}$ is a Hilbert space. Of course, it is not adequate functional analysis to study just the space l_∞ and its subspaces. Likewise subspaces of $\mathcal{L}(\mathcal{H})$

are called *concrete operator spaces,* and then subsumed in a theory of *abstract operator spaces.*

Our discussion in Section 2.6 borrows much from V. Paulsen's book. Some of our proofs are simpler because we are in finite dimensions.

Propositions 2.7.3 and 2.7.5 are due to M.-D. Choi, *Some assorted inequalities for positive linear maps on C^*-algebras,* J. Operator Theory, 4 (1980) 271–285. Propositions 2.7.6 and 2.7.8 are taken from R. Bhatia and C. Davis, *A better bound on the variance,* Am. Math. Monthly, 107 (2000) 602–608. Inequalities (2.29), (2.31) and their generalizations are important in statistics, and have been proved by many authors, often without knowledge of previous work. See the article S. W. Drury, S. Liu, C.-Y. Lu, S. Puntanen, and G. P. H. Styan, *Some comments on several matrix inequalities with applications to canonical correlations: historical background and recent developments,* Sankhyā, Series A, 64 (2002) 453–507.

The Daleckii-Krein formula was presented in Ju. L. Daleckii and S. G. Krein, *Formulas of differentiation according to a parameter of functions of Hermitian operators,* Dokl. Akad. Nauk SSSR, 76 (1951) 13–16. Infinite dimensional analogues in which the double sum in (2.38) is replaced by a double integral were proved by M. Sh. Birman and M. Z. Solomyak, *Double Stieltjes operator integrals* (English translation), *Topics in Mathematical Physics* Vol. 1, Consultant Bureau, New York, 1967.

The formula (2.41) was noted in R. Bhatia, *First and second order perturbation bounds for the operator absolute value,* Linear Algebra Appl., 208/209 (1994) 367–376. It was observed there that this equality of norms holds for several other functions that are not operator monotone. If A is positive and $f(A) = A^r$, then the equality (2.41) is true for all real numbers r other than those in $(1, \sqrt{2})$. This, somewhat mysterious, result was proved in two papers: R. Bhatia and K. B. Sinha, *Variation of real powers of positive operators,* Indiana Univ. Math. J., 43 (1994) 913–925, and R. Bhatia and J. A. R. Holbrook, *Fréchet derivatives of the power function,* ibid., 49(2000) 1155–1173. Similar equalities involving higher-order derivatives have been proved in R. Bhatia, D. Singh, and K. B. Sinha, *Differentiation of operator functions and perturbation bounds,* Commun. Math. Phys., 191 (1998) 603–611.

Chapter Three

Completely Positive Maps

For several reasons a special class of positive maps, called completely positive maps, is especially important. In Section 3.1 we study the basic properties of this class of maps. In Section 3.3 we derive some Schwarz type inequalities for this class; these are not always true for all positive maps. In Sections 3.4 and 3.5 we use general results on completely positive maps to study some important problems for matrix norms.

Let $\mathbb{M}_m(\mathbb{M}_n)$ be the space of $m \times m$ block matrices $[[A_{ij}]]$ whose i, j entry is an element of $\mathbb{M}_n = \mathbb{M}_n(\mathbb{C})$. Each linear map $\Phi : \mathbb{M}_n \to \mathbb{M}_k$ induces a linear map $\Phi_m : \mathbb{M}_m(\mathbb{M}_n) \to \mathbb{M}_m(\mathbb{M}_k)$ defined as

$$\Phi_m([[A_{ij}]]) \;=\; [[\Phi(A_{ij})]]. \tag{3.1}$$

We say that Φ is *m-positive* if the map Φ_m is positive, and Φ is *completely positive* if it is m-positive for all $m = 1, 2, \ldots$. Thus positive maps are 1-positive.

The map $\Phi(A) = A^{\mathrm{tr}}$ on $\mathbb{M}_2$ is positive but not 2-positive. To see this consider the 2×2 matrices E_{ij} whose i, j entry is one and the remaining entries are zero. Then $[[E_{ij}]]$ is positive, but $[[\Phi(E_{ij})]]$ is not.

Let $V \in \mathbb{C}^{n \times k}$, the space of $n \times k$ matrices. Then the map $\Phi(A) = V^*AV$ from $\mathbb{M}_n$ into $\mathbb{M}_k$ is completely positive. To see this note that for each m

$$[[\Phi(A_{ij})]] = (I_m \otimes V^*)[[A_{ij}]](I_m \otimes V).$$

If $V_1, \ldots, V_l \in \mathbb{C}^{n \times k}$, then

$$\Phi(A) = \sum_{j=1}^{l} V_j^* A V_j \tag{3.2}$$

is completely positive.

Let φ be any positive linear functional on $\mathbb{M}_n$. Then there exists a positive matrix X such that $\varphi(A) = \operatorname{tr} AX$ for all A. If u_j, $1 \le j \le n$, constitute an orthonormal basis for $\mathbb{C}^n$, then we have

$$\varphi(A) = \operatorname{tr} X^{1/2} A X^{1/2} = \sum_{j=1}^{n} u_j^* X^{1/2} A X^{1/2} u_j.$$

So, if we put $v_j = X^{1/2} u_j$, we have

$$\varphi(A) = \sum_{j=1}^{n} v_j^* A v_j.$$

This shows that in the special case $k = 1$, every positive linear map $\Phi : \mathbb{M}_n \to \mathbb{M}_k$ can be represented in the form (3.2) and thus is completely positive.

3.1 SOME BASIC THEOREMS

Let us fix some notations. The standard basis for $\mathbb{C}^n$ will be written as $e_j, 1 \le j \le n$. The matrix $e_i e_j^*$ will be written as E_{ij}. This is the matrix with its i, j entry equal to one and all other entries equal to zero. These matrices are called *matrix units.* The family $\{E_{ij} : 1 \le i, j \le n\}$ spans $\mathbb{M}_n$.

Our first theorem says all completely positive maps are of the form (3.2).

3.1.1 Theorem (Choi, Kraus)

Let $\Phi : \mathbb{M}_n \longrightarrow \mathbb{M}_k$ be a completely positive linear map. Then there exist $V_j \in \mathbb{C}^{n \times k}$, $1 \le j \le nk$, such that

$$\Phi(A) = \sum_{j=1}^{nk} V_j^* A V_j \quad \text{for all } A \in \mathbb{M}_n. \tag{3.3}$$

Proof. We will find V_j such that the relation (3.3) holds for all matrix units E_{rs} in $\mathbb{M}_n$. Since Φ is linear and the E_{rs} span $\mathbb{M}_n$ this is enough to prove the theorem.

We need a simple identification involving outer products of block vectors. Let $v \in \mathbb{C}^{nk}$. We think of v as a direct sum $v = x_1 \oplus \cdots \oplus x_n$, where $x_j \in \mathbb{C}^k$; or as a column vector

$$v = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \text{ where } x_j \in \mathbb{C}^k.$$

Identify this with the $k \times n$ matrix

$$V^* = [x_1, \ldots, x_n]$$

whose columns are the vectors x_j. Then note that

$$V^* E_{rs} V = [x_1, \ldots, x_n] e_r e_s^* [x_1, \ldots, x_n]^* = x_r x_s^*.$$

So, if we think of vv^* as an element of $\mathbb{M}_n(\mathbb{M}_k)$ we have

$$vv^* = [[x_r x_s^*]] \ = \ [[V^* E_{rs} V]]. \tag{3.4}$$

The matrix $[[E_{rs}]] = [[e_r e_s^*]]$ is a positive element of $\mathbb{M}_n(\mathbb{M}_n)$. So, if $\Phi : \mathbb{M}_n \to \mathbb{M}_k$ is an n-positive map, $[[\Phi(E_{rs})]]$ is a positive element of $\mathbb{M}_n(\mathbb{M}_k) = \mathbb{M}_{nk}(\mathbb{C})$.

By the spectral theorem, there exist vectors $v_j \in \mathbb{C}^{nk}$ such that

$$[[\Phi(E_{rs})]] = \sum_{j=1}^{nk} v_j v_j^* = \sum_{j=1}^{nk} [[V_j^* E_{rs} V_j]].$$

Thus for all $1 \leq r, s \leq n$

$$\Phi(E_{rs}) = \sum_{j=1}^{nk} V_j^* E_{rs} V_j,$$

as required. ■

Note that in the course of the proof we have shown that if a linear map $\Phi : \mathbb{M}_n \longrightarrow \mathbb{M}_k$ is n-positive, then it is completely positive. We have shown also that if $\Phi_n([[E_{rs}]])$ is positive, then Φ is completely positive.

The vectors v_j occurring in the proof are not unique; and so the V_j in the representation are not unique. If we impose the condition that the family $\{v_j\}$ does not contain any zero vector and all vectors in it are mutually orthogonal, then the V_j in (3.3) are unique up to unitary conjugations. The proof of this statement is left as an exercise.

The map Φ is unital if and only if $\sum V_j^* V_j = I$. Unital completely positive maps form a convex set. We state, without proof, two facts about its extreme points. The extreme points are those Φ for which the set $\{V_i^* V_j : 1 \leq i, j \leq nk\}$ is linearly independent. For such Φ, the number of terms in the representation (3.3) is at most k.

3.1.2 Theorem (The Stinespring Dilation Theorem)

Let $\Phi : \mathbb{M}_n \to \mathbb{M}_k$ be a completely positive map. Then there exist a representation

$$\Pi : \mathbb{M}_n \to \mathbb{M}_{n^2k}$$

and an operator

$$V : \mathbb{C}^k \longrightarrow \mathbb{C}^{n^2k}$$

such that $\|V\|^2 = \|\Phi(I)\|$ *and*

$$\Phi(A) = V^*\Pi(A)V.$$

Proof. The equation (3.3) can be rewritten as

$$\begin{aligned}\Phi(A) &= \sum_{j=1}^{nk} V_j^* A V_j \\ &= [V_1^*, \ldots, V_{nk}^*] \begin{bmatrix} A & & & \\ & A & & \\ & & \ddots & \\ & & & A \end{bmatrix} \begin{bmatrix} V_1 \\ \vdots \\ V_{nk} \end{bmatrix}.\end{aligned}$$

Let $V = \begin{bmatrix} V_1 \\ \vdots \\ V_{nk} \end{bmatrix}$ and $\Pi(A) = \begin{bmatrix} A & & \\ & \ddots & \\ & & A \end{bmatrix}$. ∎

Note that if Φ is unital, then $V^*V = I$. Hence V is an isometric embedding of $\mathbb{C}^k$ in $\mathbb{C}^{n^2k}$ and V^* a projection. The representation $\Pi(A) = A \otimes \cdots \otimes A$ is a direct sum of nk copies of A. This number could be smaller in several cases. The representation with the minimal number of copies is unique upto unitary conjugation.

3.1.3 Corollary

Let $\Phi : \mathbb{M}_n \to \mathbb{M}_k$ be completely positive. Then $\|\Phi\| = \|\Phi(I)\|$. (This is true, more generally, for all positive linear maps, as we saw in Chapter 2.)

Next we consider linear maps whose domain is a linear subspace $\mathcal{S} \subset \mathbb{M}_n$ and whose range is $\mathbb{M}_k$. To each element Φ of $\mathcal{L}(\mathcal{S}, \mathbb{M}_k(\mathbb{C}))$ corresponds a unique element φ of $\mathcal{L}(\mathbb{M}_k(\mathcal{S}), \mathbb{C})$. This correspondence is described as follows. Let $S_{ij}, 1 \le i, j \le k$ be elements of $\mathcal{S}$. Then

$$\varphi([[S_{ij}]]) = \frac{1}{k} \sum_{i,j=1}^{k} [\Phi(S_{ij})]_{i,j}, \tag{3.5}$$

where we use the notation $[T]_{i,j}$ for the i, j entry of a matrix T.

If $e_j, 1 \le j \le k$ is the standard basis for $\mathbb{C}^k$, and x is the vector in $\mathbb{C}^{k^2}$ given by $x = e_1 \oplus \cdots \oplus e_k$, then (3.5) can be written as

$$\varphi([[S_{ij}]]) = \frac{1}{k} \sum_{i,j=1}^{k} \langle e_i, \Phi(S_{ij}) e_j \rangle = \frac{1}{k} \langle x, [[\Phi(S_{ij})]] x \rangle. \tag{3.6}$$

In the reverse direction, suppose φ is a linear functional on $\mathbb{M}_k(\mathcal{S})$. Given an $A \in \mathcal{S}$ let $\Phi(A)$ be the element of $\mathbb{M}_k(\mathbb{C})$ whose i, j entry is

$$[\Phi(A)]_{i,j} = k\varphi(E_{ij} \otimes A), \tag{3.7}$$

where E_{ij}, $1 \le i, j \le k$, are the matrix units in $\mathbb{M}_k(\mathbb{C})$.

It is easy to see that this sets up a bijective correspondence between the spaces $\mathcal{L}(\mathcal{S}, \mathbb{M}_k(\mathbb{C}))$ and $\mathcal{L}(\mathbb{M}_k(\mathcal{S}), \mathbb{C})$. The factor $1/k$ in (3.5) ensures that Φ is unital if and only if φ is unital.

3.1.4 Theorem

Let $\mathcal{S}$ be an operator system in $\mathbb{M}_n$, and let $\Phi : \mathcal{S} \longrightarrow \mathbb{M}_k$ be a linear map. Then the following three conditions are equivalent:

(i) Φ *is completely positive.*

(ii) Φ *is k-positive.*

(iii) *The linear functional φ defined by (3.5) is positive.*

Proof. Obviously (i) $\Rightarrow$ (ii). It follows from (3.6) that (ii) $\Rightarrow$ (iii). The hard part of the proof consists of establishing the implication (iii) $\Rightarrow$ (i).

Since $\mathcal{S}$ is an operator system in $\mathbb{M}_n(\mathbb{C})$, $\mathbb{M}_k(\mathcal{S})$ is an operator system in $\mathbb{M}_k(\mathbb{M}_n) = \mathbb{M}_{kn}(\mathbb{C})$. By Krein's extension theorem (Theorem 2.6.6), the positive linear functional φ on $\mathbb{M}_k(\mathcal{S})$ has an extension $\widetilde{\varphi}$, a positive linear functional on $\mathbb{M}_k(\mathbb{M}_n)$. To this $\widetilde{\varphi}$ corresponds an element $\widetilde{\Phi}$ of $\mathcal{L}(\mathbb{M}_n(\mathbb{C}),\ \mathbb{M}_k(\mathbb{C}))$ defined via (3.7). This $\widetilde{\Phi}$ is an extension of Φ (since $\widetilde{\varphi}$ is an extension of φ). If we show $\widetilde{\Phi}$ is completely positive, it will follow that Φ is completely positive.

Let m be any positive integer. Every positive element of $\mathbb{M}_m(\mathbb{M}_n)$ can be written as a sum of matrices of the type $[[A_i^* A_j]]$ where $A_j, 1 \leq j \leq m$ are elements of $\mathbb{M}_n$. To show that $\widetilde{\Phi}$ is m-positive, it suffices to show that $[[\widetilde{\Phi}(A_i^* A_j)]]$ is positive. This is an $mk \times mk$ matrix. Let x be any vector in $\mathbb{C}^{mk}$. Write it as

$$x = x_1 \oplus \cdots \oplus x_m, \quad x_j \in \mathbb{C}^k, \quad x_j = \sum_{p=1}^{k} \xi_{jp} e_p.$$

Then

$$\begin{aligned}\langle x, [[\widetilde{\Phi}(A_i^* A_j)]]x\rangle &= \sum_{i,j=1}^{m} \langle x_i, \widetilde{\Phi}(A_i^* A_j) x_j\rangle \\ &= \sum_{i,j=1}^{m} \sum_{p,q=1}^{k} \overline{\xi}_{ip} \xi_{jq} \langle e_p, \widetilde{\Phi}(A_i^* A_j) e_q\rangle \\ &= \sum_{i,j=1}^{m} \sum_{p,q=1}^{k} \overline{\xi}_{ip} \xi_{jq} k \widetilde{\varphi}(E_{pq} \otimes A_i^* A_j),\end{aligned} \tag{3.8}$$

using (3.7). For $1 \le i \le m$ let X_i be the $k \times k$ matrix

$$X_i = \begin{bmatrix} \xi_{i1} & \xi_{i2} & \dots & \xi_{ik} \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{bmatrix}.$$

Then $[X_i^* X_j]_{p,q} = \overline{\xi}_{ip}\xi_{jq}$. In other words

$$X_i^* X_j = \sum_{p,q=1}^{k} \overline{\xi}_{ip}\xi_{jq} E_{pq}.$$

So (3.8) can be written as

$$\begin{aligned} \langle x, [[\widetilde{\Phi}(A_i^* A_j)]]x\rangle &= k \sum_{i,j=1}^{m} \widetilde{\varphi}(X_i^* X_j \otimes A_i^* A_j) \\ &= k\ \widetilde{\varphi}\left(\left(\sum_{i=1}^{m} X_i \otimes A_i\right)^* \left(\sum_{i=1}^{m} X_i \otimes A_i\right)\right). \end{aligned}$$

Since $\widetilde{\varphi}$ is positive, this expression is positive. That completes the proof. ■

In the course of the proof we have also proved the following.

3.1.5 Theorem (Arveson's Extension Theorem)

Let $\mathcal{S}$ be an operator system in $\mathbb{M}_n$ and let $\Phi : \mathcal{S} \longrightarrow \mathbb{M}_k$ be a completely positive map. Then there exists a completely positive map $\widetilde{\Phi} : \mathbb{M}_n \longrightarrow \mathbb{M}_k$ that is an extension of Φ.

Let us also record the following fact that we have proved.

3.1.6 Theorem

Let $\Phi : \mathbb{M}_n \to \mathbb{M}_k$ be a linear map. Let $m = \min(n, k)$. If Φ is m-positive, then it is completely positive.

For $l < m$, there exists a map Φ that is l-positive but not $(l+1)$-positive.

We have seen that completely positive maps have some desirable properties that positive maps did not have: they can be extended from an operator system $\mathcal{S}$ to the whole of $\mathbb{M}_n$, and they attain their norm at I for this reason (even when they have been defined only on $\mathcal{S}$). Also, there is a good characterization of completely positive maps given by (3.3). No such simple representation seems possible for positive maps. For example, one may ask whether every positive map $\Phi : \mathbb{M}_n \to \mathbb{M}_k$ is of the form

$$\Phi(A) = \sum_{i=1}^{r} V_i^* A V_i \; + \; \sum_{j=1}^{s} W_j^* A^{\text{tr}} W_j$$

for some $n \times k$ matrices V_i, W_j. For $n = k = 3$, there exist positive maps Φ that can not be represented like this.

For these reasons the notion of complete positivity seems to be more useful than that of positivity.

We remark that many of the results of this section are true in the general setting of C^*-algebras. The proofs, naturally, are more intricate in the general setting.

In view of Theorem 3.1.6, one expects that if Φ is a positive linear map from a C^*-algebra $\mathfrak{a}$ into a C^*-algebra $\mathfrak{b}$, and if either $\mathfrak{a}$ or $\mathfrak{b}$ is commutative, then Φ is completely positive. This is true.

3.2 EXERCISES

3.2.1

We have come across several positive linear maps in Chapter 2. Which of them are completely positive? What are (minimal) Stinespring dilations of these maps?

3.2.2

Every positive linear map Φ has a restricted 2-positive behaviour in the following sense:

(i) $\begin{bmatrix} A & X \\ X^* & A \end{bmatrix} \geq O \Longrightarrow \begin{bmatrix} \Phi(A) & \Phi(X) \\ \Phi(X)^* & \Phi(A) \end{bmatrix} \geq O.$

(ii) $\begin{bmatrix} A & H \\ H & B \end{bmatrix} \geq O \, (H = H^*) \Longrightarrow \begin{bmatrix} \Phi(A) & \Phi(H) \\ \Phi(H) & \Phi(B) \end{bmatrix} \geq O.$

[Hint: Use Proposition 2.7.3 and Proposition 2.7.5.]

3.2.3

Let Φ be a strictly positive linear map. Then the following three conditions are equivalent:

(i) Φ is 2-positive.

(ii) If A, B are positive matrices and X any matrix such that $B \geq X^*A^{-1}X$, then $\Phi(B) \geq \Phi(X)^*\Phi(A)^{-1}\Phi(X)$.

(iii) For every matrix X and positive A we have $\Phi(X^*A^{-1}X) \geq \Phi(X)^*\ \Phi(A)^{-1}\ \Phi(X)$.

[Compare this with Exercise 2.7.4 and Proposition 2.7.5.]

3.2.4

Let $\Phi : \mathbb{M}_3 \longrightarrow \mathbb{M}_3$ be the map defined as $\Phi(A) = 2\,(\operatorname{tr} A)\, I - A$. Then Φ is 2-positive but not 3-positive.

3.2.5

Let A and B be Hermitian matrices and suppose $A = \Phi(B)$ for some doubly stochastic map Φ on $\mathbb{M}_n$. Then there exists a completely positive doubly stochastic map Ψ such that $A = \Psi(B)$. (See Exercise 2.7.13.)

3.2.6

Let $\mathcal{S}$ be the collection of all 2×2 matrices A with $a_{11} = a_{22}$. This is an operator system in $\mathbb{M}_2$. Show that the map $\Phi(A) = A^{\mathrm{tr}}$ is completely positive on $\mathcal{S}$. What is its completely positive extension on $\mathbb{M}_2$?

3.2.7

Suppose $[[A_{ij}]]$ is a positive element of $\mathbb{M}_m(\mathbb{M}_n)$. Then each of the $m \times m$ matrices $[[\ \operatorname{tr} A_{ij}\ ]]$, $[[\ \sum\limits_{i,j} a_{ij}\ ]]$, and $[[\ \|A_{ij}\|_2^2\ ]]$ is positive.

3.3 SCHWARZ INEQUALITIES

In this section we prove some operator versions of the Schwarz inequality. Some of them are extensions of the basic inequalities for positive linear maps proved in Chapter 2.

Let μ be a probability measure on a space X and consider the Hilbert space $L_2(X,\mu)$. Let $Ef = \int f d\mu$ be the *expectation* of a function f. The *covariance* between two functions f and g in $L_2(X,\mu)$ is the quantity

$$\operatorname{cov}(f,g) = E(\bar{f}g) - \overline{Ef}Eg. \tag{3.9}$$

The *variance* of f is defined as

$$\operatorname{var}(f) = \operatorname{cov}(f,f) = E(|f|^2) - |Ef|^2. \tag{3.10}$$

(We have come across this earlier in (2.23) where we restricted ourselves to real-valued functions.) The expression (3.9) is plainly an inner product in $L_2(X,\mu)$ and the usual Schwarz inequality tells us

$$|\operatorname{cov}(f,g)|^2 \leq \operatorname{var}(f)\operatorname{var}(g). \tag{3.11}$$

This is an important, much used, inequality in statistics.

As before, replace $L_2(X,\mu)$ by $\mathbb{M}_n$ and the expectation E by a positive unital linear map Φ on $\mathbb{M}_n$. The *covariance* between two elements A and B of $\mathbb{M}_n$ (with respect to a given Φ) is defined as

$$\operatorname{cov}(A,B) = \Phi(A^*B) - \Phi(A)^*\Phi(B), \tag{3.12}$$

and *variance* of A as

$$\operatorname{var}(A) = \operatorname{cov}(A,A) = \Phi(A^*A) - \Phi(A)^*\Phi(A). \tag{3.13}$$

Kadison's inequality (2.5) says that if A is Hermitian, then $\operatorname{var}(A) \geq O$. Choi's generalization (2.6) says that this is true also when A is normal. However, with no restriction on A this is not always true. (Let $\Phi(A) = A^{\mathrm{tr}}$, and let $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$.)

If Φ is unital and 2-positive, then by Exercise 3.2.3(iii) we have

$$\Phi(A)^*\Phi(A) \leq \Phi(A^*A) \tag{3.14}$$

for all A. This says that $\operatorname{var}(A) \geq O$ for all A if Φ is 2-positive and unital. The inequality (3.14) says that

$$|\Phi(A)|^2 \leq \Phi(|A|^2). \tag{3.15}$$

The inequality $|\Phi(A)| \leq \Phi(|A|)$ is not always true even when Φ is completely positive. Let Φ be the pinching map on $\mathbb{M}_2$. If $A = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$, then $|\Phi(A)| = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ and $\Phi(|A|) = \frac{1}{\sqrt{2}}I$.

An analogue of the variance-covariance inequality (3.11) is given by the following theorem.

3.3.1 Theorem

Let Φ be a unital completely positive linear map on $\mathbb{M}_n$. Then for all A, B

$$\begin{bmatrix} \operatorname{var}(A) & \operatorname{cov}(A,B) \\ \operatorname{cov}(A,B)^* & \operatorname{var}(B) \end{bmatrix} \geq O. \tag{3.16}$$

Proof. Let V be an isometry of the space $\mathbb{C}^n$ into any $\mathbb{C}^m$. Then $V^*V = I$ and $VV^* \leq I$. From the latter condition it follows that

$$\begin{bmatrix} A^* & O \\ B^* & O \end{bmatrix} \begin{bmatrix} A & B \\ O & O \end{bmatrix} \geq \begin{bmatrix} A^* & O \\ B^* & O \end{bmatrix} \begin{bmatrix} VV^* & O \\ O & VV^* \end{bmatrix} \begin{bmatrix} A & B \\ O & O \end{bmatrix}.$$

This is the same as saying

$$\begin{bmatrix} A^*A & A^*B \\ B^*A & B^*B \end{bmatrix} \geq \begin{bmatrix} A^*VV^*A & A^*VV^*B \\ B^*VV^*A & B^*VV^*B \end{bmatrix}.$$

This inequality is preserved when we multiply both sides by the matrix $\begin{bmatrix} V^* & O \\ O & V^* \end{bmatrix}$ on the left and by $\begin{bmatrix} V & O \\ O & V \end{bmatrix}$ on the right. Thus

$$\begin{bmatrix} V^*A^*AV & V^*A^*BV \\ V^*B^*AV & V^*B^*BV \end{bmatrix} \geq \begin{bmatrix} V^*A^*VV^*AV & V^*A^*VV^*BV \\ V^*B^*VV^*AV & V^*B^*VV^*BV \end{bmatrix}.$$

This is the inequality (3.16) for the special map $\Phi(T) = V^*TV$. The general case follows from this using Theorem 3.1.2. ■

3.3.2 Remark

It is natural to wonder whether complete positivity of Φ is necessary for the inequality (3.16). It turns out that 2-positivity is not enough but 3-positivity is. Indeed, if Φ is 3-positive and unital, then from the positivity of the matrix

$$\begin{bmatrix} A^*A & A^*B & A^* \\ B^*A & B^*B & B^* \\ A & B & I \end{bmatrix} = \begin{bmatrix} A^* & O & O \\ B^* & O & O \\ I & O & O \end{bmatrix} \begin{bmatrix} A & B & I \\ O & O & O \\ O & O & O \end{bmatrix}$$

it follows that the matrix

$$\begin{bmatrix} \Phi(A^*A) & \Phi(A^*B) & \Phi(A^*) \\ \Phi(B^*A) & \Phi(B^*B) & \Phi(B^*) \\ \Phi(A) & \Phi(B) & I \end{bmatrix}$$

is positive. Hence by Theorem 1.3.3 (see Exercise 1.3.5)

$$\begin{bmatrix} \Phi(A^*A) & \Phi(A^*B) \\ \Phi(B^*A) & \Phi(B^*B) \end{bmatrix} \geq \begin{bmatrix} \Phi(A^*) & O \\ \Phi(B^*) & O \end{bmatrix} \begin{bmatrix} I & O \\ O & O \end{bmatrix} \begin{bmatrix} \Phi(A) & \Phi(B) \\ O & O \end{bmatrix}.$$

In other words,

$$\begin{bmatrix} \Phi(A^*A) & \Phi(A^*B) \\ \Phi(B^*A) & \Phi(B^*B) \end{bmatrix} \geq \begin{bmatrix} \Phi(A)^*\Phi(A) & \Phi(A)^*\Phi(B) \\ \Phi(B)^*\Phi(A) & \Phi(B)^*\Phi(B) \end{bmatrix}. \tag{3.17}$$

This is the same inequality as (3.16).

To see that this inequality is not always true for 2-positive maps, choose the map Φ on $\mathbb{M}_3$ as in Exercise 3.2.4. Let $A = E_{13}$, and $B = E_{12}$, where E_{ij} stands for the matrix whose i, j entry is one and all other entries are zero. A calculation shows that the inequality (3.17) is not true in this case.

3.3.3 Remark

If Φ is 2-positive, then for all A and B we have

$$\begin{bmatrix} \Phi(A^*A) & \Phi(A^*B) \\ \Phi(B^*A) & \Phi(B^*B) \end{bmatrix} \geq O. \tag{3.18}$$

The inequality (3.17) is a considerable strengthening of this under the additional assumption that Φ is 3-positive and unital. The inequality (3.18) is equivalent to

$$\Phi(A^*A) \geq \Phi(A^*B)\,[\Phi(B^*B)]^{-1}\,\Phi(B^*A) \tag{3.19}$$

(for 2-positive linear maps Φ). This is an operator version of the Schwarz inequality.

3.4 POSITIVE COMPLETIONS AND SCHUR PRODUCTS

A *completion problem* gives us a matrix some of whose entries are not specified, and asks us to fill in these entries in such a way that the matrix so obtained (called a *completion*) has a given property.

For example, we are given a 2×2 matrix $\begin{bmatrix} 1 & 1 \\ 1 & ? \end{bmatrix}$ with only three of its entries and are asked to choose the unknown (2,2) entry in such a way that the norm of the completed matrix is minimal among all completions. Such a completion is obtained by choosing the (2,2) entry to be -1. This is an example of a *minimal norm completion problem.*

A *positive completion problem* asks us to fill in the unspecified entries in such a way that the completed matrix is positive. Sometimes further restrictions may be placed on the completion. For example the incomplete matrix $\begin{bmatrix} ? & 1 \\ 1 & ? \end{bmatrix}$ has several positive completions: we may choose any two diagonal entries a, b such that a, b are positive and $ab \geq 1$. Among these the choice that minimises the norm of the completion is $a = b = 1$.

To facilitate further discussion, let us introduce some definitions.

A subset J of $\{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\}$ is called a *pattern.* A pattern J is called *symmetric* if

$$
\begin{gathered}
(i, i) \in J \quad \text{for} \quad 1 \leq i \leq n, \ \text{and} \\
(i, j) \in J \text{ if and only if } (j, i) \in J.
\end{gathered}
$$

We say T is a *partially defined matrix with pattern* J if the entries t_{ij} are specified for all $(i, j) \in J$. We call such a T *symmetric* if J is symmetric, t_{ii} is real for all $1 \leq i \leq n$, and $t_{ji} = \overline{t}_{ij}$ for $(i, j) \in J$.

Given a pattern J, let

$$
\mathcal{S}_J = \{A \in \mathbb{M}_n : a_{ij} = 0 \ \text{ if } \ (i, j) \notin J\}.
$$

This is a subspace of $\mathbb{M}_n$, and it is an operator system if the pattern J is symmetric.

For $T \in \mathbb{M}_n$, we use the notation S_T for the linear operator

$$
S_T(A) = T \circ A\,, \quad A \in \mathbb{M}_n
$$

and s_T for the linear functional

$$
s_T(A) = \sum_{i,j} t_{ij} a_{ij}\,, \quad A \in \mathbb{M}_n.
$$

3.4.1 Theorem

Let T be a partially defined symmetric matrix with pattern J. Then the following three conditions are equivalent:

(i) *T has a positive completion.*

(ii) *The linear map* $S_T : \mathcal{S}_J \to \mathbb{M}_n$ *is positive.*

(iii) *The linear functional* s_T on $\mathcal{S}_J$ *is positive.*

Proof. If T has a positive completion $\widetilde{T}$, then by Schur's theorem $S_{\widetilde{T}}$ is a positive map on $\mathbb{M}_n$. For $A \in \mathcal{S}_J$, $S_{\widetilde{T}}(A) = S_T(A)$. So, S_T is positive on $\mathcal{S}_J$. This proves the implication (i) ⇒ (ii). The implication (ii) ⇒ (iii) is obvious. (The sum of all entries of a positive matrix is a nonnegative number.)

(iii) ⇒ (i): Suppose s_T is positive. By Krein's extension theorem there exists a positive linear functional s on $\mathbb{M}_n$ that extends s_T. Let $\widetilde{t}_{ij} = s(E_{ij})$. Then the matrix $\widetilde{T} = \left[\left[\widetilde{t}_{ij}\right]\right]$ is a completion of T. We have for every vector x

$$\begin{aligned}\langle x, \widetilde{T}x\rangle &= \sum_{i,j} \overline{x}_i \widetilde{t}_{ij} x_j = \sum_{i,j} s(\overline{x}_i x_j E_{ij}) \\ &= s(xx^*) \geq 0.\end{aligned}$$

Thus $\widetilde{T}$ is positive. ■

For $T \in \mathbb{M}_n$ let $T^\#$ be the element of $\mathbb{M}_{2n}$ defined as $T^\# = \begin{bmatrix} I & T \\ T^* & I \end{bmatrix}$. We have seen that T is a contraction if and only if $T^\#$ is positive.

3.4.2 Proposition

Let $\mathcal{S}$ *be the operator system in* $\mathbb{M}_{2n}$ *defined as*

$$\mathcal{S} = \left\{ \begin{bmatrix} D_1 & A \\ B & D_2 \end{bmatrix} : D_1, D_2 \text{ diagonal}; A, B \in \mathbb{M}_n \right\}.$$

Then for any $T \in \mathbb{M}_n$, *the Schur multiplier* S_T *is contractive on* $\mathbb{M}_n$ *if and only if* $S_{T^\#}$ *is a positive linear map on the operator system* $\mathcal{S}$.

Proof. Suppose $S_{T^\#}$ is positive on $\mathcal{S}$. Then

$$\begin{bmatrix} I & A \\ A^* & I \end{bmatrix} \geq O \Rightarrow \begin{bmatrix} I & T \circ A \\ (T \circ A)^* & I \end{bmatrix} \geq O,$$

i.e., $\|A\| \leq 1 \Rightarrow \|T \circ A\| \leq 1$. In other words S_T is contractive on $\mathbb{M}_n$.

To prove the converse, assume $D_1, D_2 > O$, and note that

$$S_{T\#}\left(\begin{bmatrix} D_1 & A \\ A^* & D_2 \end{bmatrix}\right) = \begin{bmatrix} D_1 & T\circ A \\ (T\circ A)^* & D_2 \end{bmatrix}$$
$$= \begin{bmatrix} D_1^{1/2} & O \\ O & D_2^{1/2} \end{bmatrix} \begin{bmatrix} I & D_1^{-1/2}(T\circ A)D_2^{-1/2} \\ D_2^{-1/2}(T\circ A)^* D_1^{-1/2} & I \end{bmatrix} \begin{bmatrix} D_1^{1/2} & O \\ O & D_2^{1/2} \end{bmatrix}$$
$$\sim \begin{bmatrix} I & D_1^{-1/2}(T\circ A)D_2^{-1/2} \\ D_2^{-1/2}(T\circ A)^* D_1^{-1/2} & I \end{bmatrix}$$
$$= \begin{bmatrix} I & T\circ (D_1^{-1/2}AD_2^{-1/2}) \\ (T\circ D_1^{-1/2}AD_2^{-1/2})^* & I \end{bmatrix}.$$

If S_T is contractive on $\mathbb{M}_n$, then

$$\|D_1^{-1/2}AD_2^{-1/2}\| \le 1 \Rightarrow \|T\circ D_1^{-1/2}AD_2^{-1/2}\| \le 1,$$

i.e.,

$$\begin{bmatrix} D_1 & A \\ A^* & D_2 \end{bmatrix} \ge O \Rightarrow \begin{bmatrix} I & T\circ (D_1^{-1/2}AD_2^{-1/2}) \\ (T\circ D_1^{-1/2}AD_2^{-1/2})^* & I \end{bmatrix} \ge O.$$

We have seen above that the last matrix is congruent to $S_{T\#}\left(\left[\begin{smallmatrix} D_1 & A \\ A^* & D_2 \end{smallmatrix}\right]\right)$. This shows that $S_{T\#}$ is positive on $\mathcal{S}$. ■

We can prove now the main theorem of this section.

3.4.3 Theorem (Haagerup's Theorem)

Let $T \in \mathbb{M}_n$. Then the following four conditions are equivalent:

(i) *S_T is contractive; i.e., $\|T\circ A\| \le \|A\|$ for all A.*

(ii) *There exist vectors $v_j, w_j, 1 \le j \le n$, all with their norms ≤ 1, such that $t_{ij} = v_i^* w_j$.*

(iii) *There exist positive matrices R_1, R_2 with $\operatorname{diag} R_1 \le I$, $\operatorname{diag} R_2 \le I$ and such that $\left[\begin{smallmatrix} R_1 & T \\ T^* & R_2 \end{smallmatrix}\right]$ is positive.*

(iv) *T can be factored as $T = V^*W$ with $\|V\|_c \le 1$, $\|W\|_c \le 1$. (The symbol $\|Y\|_c$ stands for the maximum of the Euclidean norms of the columns of Y.)*

Proof. Let S_T be contractive. Then, by Proposition 3.4.2, $S_{T^\#}$ is a positive operator on the operator system $\mathcal{S} \subset \mathbb{M}_{2n}$. By Theorem 3.4.1, $T^\#$ has a positive completion. (Think of the off-diagonal entries of the two diagonal blocks as unspecified.) Call this completion P. It has a Cholesky factoring $P = \triangle^*\triangle$ where $\triangle$ is an upper triangular $2n \times 2n$ matrix. Write $\triangle = \begin{bmatrix} V & W \\ O & X \end{bmatrix}$. Then

$$P = \begin{bmatrix} V^*V & V^*W \\ W^*V & W^*W + X^*X \end{bmatrix}.$$

Let $v_j, w_j, 1 \le j \le n$ be the columns of V, W, respectively. Since P is a completion of $T^\#$, we have $T = V^*W$; i.e., $t_{ij} = v_i^* w_j$. Since $\operatorname{diag}(V^*V) = I$, we have $\|v_j\| = 1$. Since $\operatorname{diag}(W^*W + X^*X) = I$, we have $\|w_j\| \le 1$. This proves the implication (i) $\Rightarrow$ (ii).

The condition (ii) can be expressed by saying $T = V^*W$, where $\operatorname{diag}(V^*V) \le I$ and $\operatorname{diag}(W^*W) \le I$. Since

$$\begin{bmatrix} V^*V & V^*W \\ W^*V & W^*W \end{bmatrix} = \begin{bmatrix} V^* & O \\ W^* & O \end{bmatrix} \begin{bmatrix} V & W \\ O & O \end{bmatrix} \ge O,$$

this shows that the statement (ii) implies (iii). Clearly (iv) is another way of stating (ii).

To complete the proof we show that (iii) $\Rightarrow$ (i). Let $A \in \mathbb{M}_n$, $\|A\| \le 1$. This implies $\begin{bmatrix} I & A \\ A^* & I \end{bmatrix} \ge O$. Then the condition (iii) leads to the inequality

$$\begin{aligned} \begin{bmatrix} I & T \circ A \\ (T \circ A)^* & I \end{bmatrix} &\ge \begin{bmatrix} R_1 \circ I & T \circ A \\ (T \circ A)^* & R_2 \circ I \end{bmatrix} \\ &= \begin{bmatrix} R_1 & T \\ T^* & R_2 \end{bmatrix} \circ \begin{bmatrix} I & A \\ A^* & I \end{bmatrix} \\ &\ge O. \end{aligned}$$

But this implies $\|T \circ A\| \le 1$. In other words S_T is contractive. ■

3.4.4 Corollary

For every T in $\mathbb{M}_n$, we have $\|S_T\| = \min \{\|V\|_c \, \|W\|_c : T = V^\star W\}$.

3.5 THE NUMERICAL RADIUS

The *numerical range* of an operator A is the set of complex numbers

$$W(A) = \Big\{ \langle x, Ax \rangle : \|x\| = 1 \Big\},$$

and the *numerical radius* is the number

$$w(A) = \sup_{\|x\|=1} |\langle x, Ax \rangle| = \sup \Big\{ |z| : z \in W(A) \Big\}.$$

It is known that the set $W(A)$ is convex, and $w(\cdot)$ defines a norm. We have

$$w(A) \leq \|A\| \leq 2w(A) \text{ for all } A.$$

Some properties of w are summarised below. It is not difficult to prove them.

(i) $w(UAU^*) = w(A)$ for all A, and unitary U.

(ii) If A is diagonal, then $w(A) = \max |a_{ii}|$.

(iii) More generally,

$$w(A_1 \oplus \cdots \oplus A_k) = \max_{1 \leq j \leq k} w(A_j).$$

(iv) $w(A) = \|A\|$ if (but not only if) A is normal.

(v) w is not submultiplicative: the inequality $w(AB) \leq w(A)w(B)$ is not always true for 2×2 matrices.

(vi) Even the weaker inequality $w(AB) \leq \|A\| w(B)$ is not always true for 2×2 matrices.

(vii) The inequality $w(A \otimes B) \leq w(A)w(B)$ is not always true for 2×2 matrices A, B.

(viii) However, we do have $w(A \otimes B) \leq \|A\| w(B)$ for square matrices A, B of any size.

(Proof: It is enough to prove this when $\|A\| = 1$. Then $A = \frac{1}{2}(U + V)$ where U, V are unitary. So it is enough to prove that $w(U \otimes B) \leq w(B)$ if U is unitary. Choose an orthonormal basis in which U is diagonal, and use (iii).)

(ix) If $w(A) \leq 1$, then $I \pm \text{Re}A \geq O$.
($|\text{Re}\langle x, Ax\rangle| \leq |\langle x, Ax\rangle| \leq \langle x, x\rangle$ for all x.)

(x) The inequality $w(AB) \leq w(A)w(B)$ may not hold even when A, B commute. Let

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

Then $w(A) < 1, w(A^2) = w(A^3) = 1/2$. So $w(A^3) > w(A)w(A^2)$ in this case.

Proposition 1.3.1 characterizes operators A with $\|A\| \leq 1$ in terms of positivity of certain 2×2 block matrices. A similar theorem for operators A with $w(A) \leq 1$ is given below.

3.5.1 Theorem (Ando)

Let $A \in \mathbb{M}_n$. Then $w(A) \leq 1$ if and only if there exists a Hermitian matrix H such that $\begin{bmatrix} I+H & A \\ A^* & I-H \end{bmatrix}$ is positive.

Proof. If $\begin{bmatrix} I+H & A \\ A^* & I-H \end{bmatrix} \geq O$, then there exists an operator K with $\|K\| \leq 1$ such that $A = (I+H)^{1/2}K(I-H)^{1/2}$. So, for every vector x

$$\begin{aligned} |\langle x, Ax\rangle| &= |\langle x, (I+H)^{1/2}K(I-H)^{1/2}x\rangle| \\ &\leq \|(I+H)^{1/2}x\| \, \|(I-H)^{1/2}x\| \\ &\leq \frac{1}{2}\left(\|(I+H)^{1/2}x\|^2 + \|(I-H)^{1/2}x\|^2\right) \\ &= \frac{1}{2}\left(\langle x, (I+H)x\rangle + \langle x, (I-H)x\rangle\right) \\ &= \|x\|^2. \end{aligned}$$

This shows that $w(A) \leq 1$.

The proof of the other half of the theorem is longer. Let A be an operator with $w(A) \leq 1$. Let $\mathcal{S}$ be the collection of 2×2 matrices

$\begin{bmatrix} x & y \\ z & x \end{bmatrix}$ where x, y, z are complex numbers. Then $\mathcal{S}$ is an operator system. Let $\Phi : \mathcal{S} \to \mathbb{M}_n$ be the unital linear map defined as

$$\Phi\left(\begin{bmatrix} x & y \\ z & x \end{bmatrix}\right) = xI + \frac{1}{2}(yA + zA^*).$$

It follows from property (ix) listed at the beginning of the section that Φ is positive. We claim it is completely positive. Let m be any positive integer. We want to show that if the $m \times m$ block matrix with the 2×2 block $\begin{bmatrix} x_{ij} & y_{ij} \\ z_{ij} & x_{ij} \end{bmatrix}$ as its i, j entry is positive, then the $m \times m$ block matrix with the $n \times n$ block $x_{ij}I + \frac{1}{2}(y_{ij}A + z_{ij}A^*)$ as its i, j entry is also positive. Applying permutation similarity the first matrix can be converted to a matrix of the form $\begin{bmatrix} X & Y \\ Z & X \end{bmatrix}$ where X, Y, Z are $m \times m$ matrices. If this is positive, then we have $Z = Y^*$, and our claim is that

$$\begin{bmatrix} X & Y \\ Y^* & X \end{bmatrix} \geq O \Rightarrow X \otimes I_n + \frac{1}{2}(Y \otimes A + Y^* \otimes A^*) \geq O.$$

We can apply a congruence, and replace the matrices X by I and Y by $X^{-1/2}YX^{-1/2}$, respectively. Thus we need to show that

$$\begin{bmatrix} I_m & Y \\ Y^* & I_m \end{bmatrix} \geq O \Rightarrow I_m \otimes I_n + \frac{1}{2}(Y \otimes A + Y^* \otimes A^*) \geq O.$$

The hypothesis here is (equivalent to) $\|Y\| \leq 1$. By property (viii) this implies $w(Y \otimes A) \leq w(A) \leq 1$. So the conclusion follows from property (ix).

We have shown that Φ is completely positive on $\mathcal{S}$. By Arveson's theorem Φ can be extended to a completely positive map $\widetilde{\Phi} : \mathbb{M}_2 \to \mathbb{M}_n$.

Let $E_{ij}, 1 \leq i, j \leq 2$ be the matrix units in $\mathbb{M}_2$. Then the matrix $[[\widetilde{\Phi}(E_{ij})]]$ is positive. Thus, in particular, $\widetilde{\Phi}(E_{11})$ and $\widetilde{\Phi}(E_{22})$ are positive, and their sum is I since $\widetilde{\Phi}$ is unital.

Put $H = \widetilde{\Phi}(E_{11}) - \widetilde{\Phi}(E_{22})$. Then H is Hermitian, and

$$\widetilde{\Phi}(E_{11}) = \frac{I + H}{2}, \ \widetilde{\Phi}(E_{22}) = \frac{I - H}{2}.$$

Since $\widetilde{\Phi}$ is an extension of Φ, we have

$$\widetilde{\Phi}(E_{12}) = \frac{1}{2}A, \ \widetilde{\Phi}(E_{21}) = \frac{1}{2}A^*.$$

Thus

$$[[\widetilde{\Phi}(E_{ij})]] = \frac{1}{2}\begin{bmatrix} I+H & A \\ A^* & I-H \end{bmatrix},$$

and this matrix is positive. ■

3.5.2 Corollary

For every A and $k = 1, 2, \dots$

$$w(A^k) \le w(A)^k. \tag{3.20}$$

Proof. It is enough to show that if $w(A) \le 1$, then $w(A^k) \le 1$. Let $w(A) \le 1$. By Ando's theorem, there exists a Hermitian matrix H such that

$$\begin{bmatrix} I+H & A \\ A^* & I-H \end{bmatrix} \ge O.$$

Hence, there exists a contraction K such that

$$A = (I+H)^{1/2} K (I-H)^{1/2}.$$

Then

$$\begin{aligned} A^k &= (I+H)^{1/2} K [(I-H^2)^{1/2} K]^{k-1} (I-H)^{1/2} \\ &= (I+H)^{1/2} L (I-H)^{1/2}, \end{aligned}$$

where $L = K[(I-H^2)^{1/2}K]^{k-1}$ is a contraction. But this implies that

$$\begin{bmatrix} I+H & A^k \\ A^{*k} & I-H \end{bmatrix} \ge O.$$

So, by Ando's Theorem $w(A^k) \leq 1$. ■

The inequality (3.20) is called the *power inequality* for the numerical radius.

Ando and Okubo have proved an analogue of Haagerup's theorem for the norm of the Schur product with respect to the numerical radius. We state it without proof.

3.5.3 Theorem (Ando-Okubo)

Let T be any matrix. Then the following statements are equivalent:

(i) $w(T \circ A) \leq 1$ *whenever* $w(A) \leq 1$.

(ii) *There exists a positive matrix R with* $\text{diag}R \leq I$ *such that*

$$\begin{bmatrix} R & T \\ T^* & R \end{bmatrix} \geq O.$$

3.6 SUPPLEMENTARY RESULTS AND EXERCISES

The Schwarz inequality, in its various forms, is the most important inequality in analysis. The first few remarks in this section supplement the discussion in Section 3.3.

Let A be an $n \times k$ matrix and B an $n \times l$ matrix of rank l. The matrix

$$\begin{bmatrix} A^*A & A^*B \\ B^*A & B^*B \end{bmatrix}$$

is positive. This is equivalent to the assertion

$$A^*A \geq A^*B(B^*B)^{-1}B^*A. \tag{3.21}$$

This is a matrix version of the Schwarz inequality. It can be proved in another way as follows. The matrix $B(B^*B)^{-1}B^*$ is idempotent and Hermitian. Hence $I \geq B(B^*B)^{-1}B^*$ and (3.21) follows immediately. The inequality (3.19) is an extension of (3.21).

Let A be a positive operator and let x, y be any two vectors. From the Schwarz inequality we get

$$|\langle x, Ay\rangle|^2 \leq \langle x, Ax\rangle\langle y, Ay\rangle. \tag{3.22}$$

An operator version of this in the spirit of (3.19) can be obtained as follows. For any two operators X and Y we have

$$\begin{bmatrix} X^*AX & X^*AY \\ Y^*AX & Y^*AY \end{bmatrix} = \begin{bmatrix} X^* & O \\ O & Y^* \end{bmatrix} \begin{bmatrix} A & A \\ A & A \end{bmatrix} \begin{bmatrix} X & O \\ O & Y \end{bmatrix} \geq O.$$

So, if Φ is a 2-positive linear map, then

$$\begin{bmatrix} \Phi(X^*AX) & \Phi(X^*AY) \\ \Phi(Y^*AX) & \Phi(Y^*AY) \end{bmatrix} \geq O,$$

or, equivalently,

$$\Phi(X^*AY)\,[\Phi(Y^*AY)]^{-1}\,\Phi(Y^*AX) \leq \Phi(X^*AX). \tag{3.23}$$

This is an operator version of (3.22).

There is a considerable strengthening of the inequality (3.22) in the special case when x is orthogonal to y. This says that if A is a positive operator with $mI \leq A \leq MI$, and $x \perp y$, then

$$|\langle x, Ay\rangle|^2 \leq \left(\frac{M-m}{M+m}\right)^2 \langle x, Ax\rangle\langle y, Ay\rangle. \tag{3.24}$$

This is called *Wielandt's inequality.* The following theorem gives an operator version.

3.6.1 Theorem

Let A be a positive element of $\mathbb{M}_n$ with $mI \leq A \leq MI$. Let X, Y be two mutually orthogonal projection operators in $\mathbb{C}^n$. Then for every 2-positive linear map Φ on $\mathbb{M}_n$ we have

$$\Phi(X^*AY)\,[\Phi(Y^*AY)]^{-1}\,\Phi(Y^*AX) \leq \left(\frac{M-m}{M+m}\right)^2 \Phi(X^*AX). \tag{3.25}$$

Proof. First assume that $X \oplus Y = I$. With respect to this decomposition, let A have the block form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

By Exercise 1.5.7

$$A^{-1} = \begin{bmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & \star \\ \star & \star \end{bmatrix}.$$

Apply Proposition 2.7.8 with Φ as the pinching map. This shows

$$\left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1} \leq \frac{(M+m)^2}{4Mm}A_{11}^{-1}.$$

Taking inverses changes the direction of this inequality, and then rearranging terms we get

$$A_{12}A_{22}^{-1}A_{21} \leq \left(\frac{M-m}{M+m}\right)^2 A_{11}.$$

This is the inequality (3.25) in the special case when Φ is the identity map. A minor argument shows that the assumption $X \oplus Y = I$ can be dropped.

Let $\alpha = (M-m)/(M+m)$. The inequality we have just proved is equivalent to the statement

$$\begin{bmatrix} \alpha X^*AX & X^*AY \\ Y^*AX & Y^*AY \end{bmatrix} \geq O.$$

This implies that the inequality (3.25) holds for every 2-positive linear map Φ. ■

We say that a complex function f on $\mathbb{M}_n$ is in the Lieb class $\mathcal{L}$ if $f(A) \geq 0$ whenever $A \geq O$, and $|f(X)|^2 \leq f(A)f(B)$ whenever $\begin{bmatrix} A & X \\ X^* & B \end{bmatrix} \geq O$. Several examples of such functions are given in MA (pages 268–270). We have come across several interesting 2×2 block matrices that are positive. Many Schwarz type inequalities for functions in the class $\mathcal{L}$ can be obtained from these block matrices.

The next few results concern maps associated with pinchings and their norms.

Let $\mathcal{D}(A)$ be the diagonal part of a matrix:

$$\mathcal{D}(A) = \operatorname{diag}(A) = \sum_{j=1}^{n} P_jAP_j, \tag{3.26}$$

where $P_j = e_je_j^*$ is the orthogonal projection onto the one-dimensional space spanned by the vector e_j. This is a special case of the pinching operation $\mathcal{C}$ introduced in Example 2.2.1 (vii). Since $\sum P_j = I$ and $P_j \geq O$, we think of the sum (3.26) as a noncommutative convex combination. There is another interesting way of expressing for $\mathcal{D}(A)$. Let $\omega = e^{2\pi i/n}$ and let U be the diagonal unitary matrix

$$U = \mathrm{diag}(1, \omega, \ldots, \omega^{n-1}). \tag{3.27}$$

Then

$$\mathcal{D}(A) = \frac{1}{n} \sum_{k=0}^{n-1} U^{*k} A U^{k}. \tag{3.28}$$

(The sum on the right-hand side is the Schur product of A by a matrix whose i, j entry is

$$\sum_{k=0}^{n-1} \omega^{k(j-i)} = n\delta_{ij}.)$$

This idea can be generalized.

3.6.2 Exercise

Partition $n \times n$ matrices into an $r \times r$ block form in which the diagonal blocks are square matrices of dimension $d_1, \ldots, d_r$. Let $\mathcal{C}$ be the pinching operation sending the block matrix $A = [[A_{ij}]]$ to the block diagonal matrix $\mathcal{C}(A) = \mathrm{diag}(A_{11}, \ldots, A_{rr})$. Let $\omega = e^{2\pi i/r}$ and let V be the diagonal unitary matrix

$$V = \mathrm{diag}(I_1, \omega I_2, \ldots, \omega^{r-1} I_r)$$

where I_j is the identity matrix of size d_j. Show that

$$\mathcal{C}(A) = \frac{1}{r} \sum_{k=0}^{r-1} V^{*k} A V^{k}. \tag{3.29}$$

3.6.3 Exercise

Let J be a pattern and let $\mathcal{J}$ be the map on $\mathbb{M}_n$ induced by J as follows. The i, j entry of $\mathcal{J}(A)$ is a_{ij} for all $(i, j) \in J$ and is zero otherwise. Suppose J is an equivalence relation on $\{1, 2, \ldots, n\}$ and has r equivalence classes. Show that

$$\mathcal{J}(A) = \frac{1}{r}\sum_{k=0}^{r-1} W^{*k} A W^k, \tag{3.30}$$

where W is a diagonal unitary matrix. Conversely, show that if $\mathcal{J}$ can be represented as

$$\mathcal{J}(A) = \sum_{k=0}^{r-1} \lambda_k U_k^* A U_k, \tag{3.31}$$

where U_j are unitary matrices and λ_j are positive numbers with $\sum \lambda_j = 1$, then J is an equivalence relation with r equivalence classes. It is not possible to represent $\mathcal{J}$ as a convex combination of unitary transforms as in (3.31) with fewer than r terms.

3.6.4 Exercise

Let V be the permutation matrix

$$V = \begin{bmatrix} 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ . & . & . & . & . \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}. \tag{3.32}$$

Show that

$$\mathcal{D}\left(\sum_{k=0}^{n-1} V^{*k} A V^k\right) = \frac{\operatorname{tr} A}{n} I. \tag{3.33}$$

Find n^2 unitary matrices W_j such that

$$\frac{\operatorname{tr} A}{n} I = \sum_{j=1}^{n^2} W_j^* A W_j \text{ for all } A. \tag{3.34}$$

This gives a representation of the linear map $\mathcal{T}(A) = \frac{\operatorname{tr} A}{n} I$ from $\mathbb{M}_n$ into scalar matrices.

It is of some interest to consider what is left of a matrix after the diagonal part is removed. Let

$$\mathcal{O}(A) = A - \mathcal{D}(A) \tag{3.35}$$

be the *off-diagonal* part of A. Using (3.28) we can write

$$\mathcal{O}(A) = \left(1 - \frac{1}{n}\right) A + \frac{1}{n} \sum_{k=1}^{n-1} U^{*k} A U^k.$$

From this we get

$$\|\mathcal{O}(A)\| \le 2\left(1 - \frac{1}{n}\right) \|A\|. \tag{3.36}$$

This inequality is sharp. To see this choose $A = E - \frac{n}{2} I$, where E is the matrix all of whose entries are equal to one.

3.6.5 Exercise

Let $B = E - I$. We have just seen that the Schur multiplier norm

$$\|S_B\| = 2\left(1 - \frac{1}{n}\right). \tag{3.37}$$

Find an alternate proof of this using Theorem 3.4.3.

3.6.6 Exercise

Use Exercise 3.6.3 to show that

$$\|A - \mathcal{T}(A)\| \le 2\left(1 - \frac{1}{n^2}\right) \|A\| \quad \text{for all } A.$$

This inequality can be improved:

(i) Every matrix is unitarily similar to one with constant diagonal entries. [Prove this by induction, with the observation that $\frac{\operatorname{tr} A}{n} = \langle x, Ax \rangle$ for some unit vector x.]

(ii) Thus, in some orthonormal basis, removing $\mathcal{T}(A)$ has the same effect as removing $\mathcal{D}(A)$ from A. Thus

$$\|A - \mathcal{T}(A)\| \leq 2\left(1 - \frac{1}{n}\right)\|A\| \ \text{ for all } A, \tag{3.38}$$

and this inequality is sharp.

3.6.7 Exercise

The Schur multiplier norm is multiplicative over tensor products; i.e.,

$$\|S_{A\otimes B}\| = \|S_A\| \ \|S_B\| \ \text{ for all } A, B.$$

3.6.8 Exercise

Let $B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Show, using Theorem 3.4.3 and otherwise, that

$$\|S_B\| = \frac{2}{\sqrt{3}}.$$

Let $\triangle_n$ be the *triangular truncation operator* taking every $n \times n$ matrix to its upper triangular part. Then we have $\|\triangle_2\| = 2/\sqrt{3}$. Try to find $\|\triangle_3\|$.

3.6.9 Exercise

Fill in the details in the following proof of the power inequality (3.20).

(i) If a is a complex number, then $|a| \leq 1$ if and only if $\mathrm{Re}(1 - za) \geq 0$ for all z with $|z| < 1$.

(ii) $w(A) \leq 1$ if and only if $\mathrm{Re}(I - zA) \geq O$ for $|z| < 1$.

(iii) $w(A) \leq 1$ if and only if $\mathrm{Re}((I - zA)^{-1}) \geq O$ for $|z| < 1$.

(iv) Let $\omega = e^{2\pi i/k}$. Prove the identity

$$\frac{1}{1 - z^k} = \frac{1}{k}\sum_{j=0}^{k-1} \frac{1}{1 - \omega^j z} \qquad \text{if } \ z^k \neq 1.$$

(v) If $w(A) \le 1$, then

$$(I - z^k A^k)^{-1} = \frac{1}{k}\sum_{j=0}^{k-1}(I - \omega^j z A)^{-1}, \quad \text{for } |z| < 1.$$

(vi) Assume $w(A) \le 1$. Use (v) and (iii) to conclude that $w(A^k) \le 1$.

By Exercise 3.2.7, if $[[A_{ij}]]$ is a positive element of $\mathbb{M}_m(\mathbb{M}_n)$, then the $m \times m$ matrices $[[\ \mathrm{tr}\, A_{ij}\ ]]$ and $[[\ \|A_{ij}\|_2^2\ ]]$ are positive. Matricial curiosity should make us wonder whether this remains true when tr is replaced by other matrix functions like det, and the norm $\|\cdot\|_2$ is replaced by the norm $\|\cdot\|$.

For the sake of economy, in the following discussion we use (temporarily) the terms positive, m-positive, and completely positive to encompass *nonlinear* maps as well. Thus we say a map $\Phi : \mathbb{M}_n \to \mathbb{M}_k$ is positive if $\Phi(A) \ge O$ whenever $A \ge O$, and completely positive if $[[\Phi(A_{ij})]]$ is positive whenever a block matrix $[[A_{ij}]]$ is positive. For example, $\det(A)$ is a positive (nonlinear) function, and we have observed that $\Phi(A) = \|A\|_2^2$ is a completely positive (nonlinear) function. In Chapter 1 we noted that a function $\varphi : \mathbb{C} \to \mathbb{C}$ is completely positive if and only if it can be expressed in the form (1.40).

3.6.10 Proposition

Let $\varphi(A) = \|A\|^2$. Then φ is 2-positive but not 3-positive.

Proof. The 2-positivity is an easy consequence of Proposition 1.3.2. The failure of φ to be 3-positive is illustrated by the following example in $\mathbb{M}_3(\mathbb{M}_2)$. Let

$$X = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad Y = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad Z = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

Since X, Y and Z are positive, so is the matrix

$$A = \begin{bmatrix} X & X & X \\ X & X & X \\ X & X & X \end{bmatrix} + \begin{bmatrix} Y & Y & O \\ Y & Y & O \\ O & O & O \end{bmatrix} + \begin{bmatrix} O & O & O \\ O & Z & Z \\ O & Z & Z \end{bmatrix}.$$

If we write A as $[[A_{ij}]]$ where A_{ij}, $1 \le i, j \le 3$ are 2×2 matrices, and replace each A_{ij} by $\|A_{ij}\|^2$ we obtain the matrix

$$\begin{bmatrix} \alpha & \alpha & 1 \\ \alpha & 9 & \alpha \\ 1 & \alpha & \alpha \end{bmatrix}, \quad \alpha = \frac{7+\sqrt{45}}{2}.$$

This matrix is not positive as its determinant is negative. ■

3.6.11 Exercise

Let $\Phi : \mathbb{M}_2 \to \mathbb{M}_2$ be the map defined as $\Phi(X) = |X|^2 = X^*X$. Use the example in Exercise 1.6.6 to show that Φ is not two-positive.

3.6.12 Exercise

Let $\otimes^k A = A \otimes \cdots \otimes A$ be the k-fold tensor power of A. Let $A = [[A_{ij}]]$ be an element of $\mathbb{M}_m(\mathbb{M}_n)$. Then $\otimes^k A$ is a matrix of size $(mn)^k$ whereas $[[\otimes^k A_{ij}]]$ is a matrix of size mn^k. Show that the latter is a principal submatrix of the former. Use this observation to conclude that $\otimes^k$ is a completely positive map from $\mathbb{M}_n$ to $\mathbb{M}_{n^k}$.

3.6.13 Exercise

For $1 \le k \le n$ let $\wedge^k A$ be the kth antisymmetric tensor power of an $n \times n$ matrix A. Show that $\wedge^k$ is a completely positive map from $\mathbb{M}_n$ into $\mathbb{M}_{\binom{n}{k}}$. If

$$t^n - c_1(A)t^{n-1} + c_2(A)t^{n-2} - \cdots + (-1)^n c_n(A)$$

is the characteristic polynomial of A, then $c_k(A) = \operatorname{tr} \wedge^k A$. Hence each c_k is a completely positive functional. In particular, det is completely positive.

Similar considerations apply to other "symmetry classes" of tensors and the associated "Schur functions." Thus, for example, the permanent function is completely positive.

3.6.14 Exercise

Let $\Phi : \mathbb{M}_n \to \mathbb{M}_k$ be any 4-positive map. Let X, Y, Z be positive elements of $\mathbb{M}_n$ and let

$$A = \begin{bmatrix} X+Y & X+Y & X & X \\ X+Y & X+Y+Z & X+Z & X \\ X & X+Z & X+Z & X \\ X & X & X & X \end{bmatrix}.$$

Then $A = [[A_{ij}]]$ *is positive. Let* $X = [I, -I, I, -I]$. *Consider the product* $X[[\Phi(A_{ij})]]X^*$ *and conclude that*

$$\Phi(Y + X) + \Phi(X + Z) \leq \Phi(X) + \Phi(X + Y + Z). \tag{3.39}$$

Inequalities of the form (3.39) occur in other contexts. For example, if P, Q and R are (rectangular) matrices and the product PQR is defined, then the *Frobenius inequality* is the relation between ranks:

$$\operatorname{rk}(PQ) + \operatorname{rk}(QR) \leq \operatorname{rk}(Q) + \operatorname{rk}(PQR).$$

The inequality (4.49) in Chapter 4 is another one with a similar structure.

3.7 NOTES AND REFERENCES

The theory of completely positive maps has been developed by operator algebraists and mathematical physicists over the last four decades.

The two major results of Section 3.1, the theorems of Stinespring and Arveson, hold in much more generality. We have given their baby versions by staying in finite dimensions.

Stinespring's theorem was proved in W. F. Stinespring, *Positive functions on C^*-algebras,* Proc. Amer. Math. Soc., 6 (1955) 211–216. To put it in context, it is helpful to recall an earlier theorem due to M. A. Naimark.

Let $(X, \mathcal{S})$ be a compact Hausdorff space with its Borel σ-algebra $\mathcal{S}$, and let $\mathcal{P}(\mathcal{H})$ be the collection of orthogonal projections in a Hilbert space $\mathcal{H}$. A *projection-valued measure* is a map $S \longmapsto P(S)$ from $\mathcal{S}$ into $\mathcal{P}(\mathcal{H})$ that is countably additive: if $\{S_i\}$ is a countable collection of disjoint sets, then

$$\langle P\Big(\cup_{i=1}^{\infty} S_i\Big)x, y\rangle = \sum_{i=1}^{\infty}\langle P(S_i)x, y\rangle$$

for all x and y in $\mathcal{H}$. The spectral theorem says that if A is a bounded self-adjoint operator on $\mathcal{H}$, then there exists a projection-valued measure on $[-\|A\|, \|A\|]$ taking values in $\mathcal{P}(\mathcal{H})$, and with respect to this measure A can be written as the integral $A = \int \lambda dP(\lambda)$.

Instead of projection-valued measures we may consider an *operator-valued measure.* This assigns to each set S an element $E(S)$ of $\mathcal{L}(\mathcal{H})$, the map is countably additive, and $\sup\{\|E(S)\| : S \in \mathcal{S}\} < \infty$. Such a measure gives rise to a complex measure

$$\mu_{x,y}(\mathcal{S}) = \langle E(S)x, y\rangle \tag{3.40}$$

for each pair x, y in $\mathcal{H}$. This in turn gives a bounded linear map Φ from the space $\mathrm{C}(X)$ into $\mathcal{L}(\mathcal{H})$ via

$$\langle \Phi(f)x, y\rangle = \int f d\mu_{x,y}. \tag{3.41}$$

This process can be reversed. Given a bounded linear map $\Phi : \mathrm{C}(X) \to \mathcal{L}(\mathcal{H})$ we can construct complex measures $\mu_{x,y}$ via (3.40) and then an operator-valued measure E via (3.39). If $E(S)$ is a positive operator for all S, we say the measure E is positive.

Naimark's theorem says that every positive operator-valued measure can be dilated to a projection-valued measure. More precisely, if E is a positive $\mathcal{L}(\mathcal{H})$-valued measure on $(X, \mathcal{S})$, then there exist a Hilbert space $\mathcal{K}$, a bounded linear map $V : \mathcal{H} \to \mathcal{K}$, and a $\mathcal{P}(\mathcal{H})$-valued measure P such that

$$E(S) = V^*P(S)V \quad \text{for all } S \text{ in } \mathcal{S}.$$

The point of the theorem is that by dilating to the space $\mathcal{K}$ we have replaced the operator-valued measure E by the projection-valued measure P which is nicer in two senses: it is more familiar because of its connections with the spectral theorem and the associated map Φ is now a $*$-homomorphism of $\mathrm{C}(X)$.

The Stinespring theorem is a generalization of Naimark's theorem in which the commutative algebra $\mathrm{C}(X)$ is replaced by a unital C^*-algebra. The theorem in its full generality says the following. If Φ is a completely positive map from a unital C^*-algebra $\mathfrak{a}$ into $\mathcal{L}(\mathcal{H})$, then there exist a Hilbert space $\mathcal{K}$, a unital $*$-homomorphism (i.e., a representation) $\Pi : \mathfrak{a} \to \mathcal{L}(\mathcal{K})$, and a bounded linear operator $V : \mathcal{H} \to \mathcal{K}$ with $\|V\|^2 = \|\Phi(I)\|$ such that

$$\Phi(A) = V^*\Pi(A)V \quad \text{for all } A \in \mathfrak{a}.$$

A "minimal" Stinespring dilation (in which $\mathcal{K}$ is a smallest possible space) is unique up to unitary equivalence.

The term *completely positive* was introduced in this paper of Stinespring. The theory of positive and completely positive maps was vastly expanded in the hugely influential papers by W. B. Arveson, *Subalgebras of C^*-algebras, I, II,* Acta Math. 123 (1969) 141–224 and 128 (1972) 271–308. In the general version of Theorem 3.1.5 the space $\mathbb{M}_n$ is replaced by an arbitrary C^*-algebra $\mathfrak{a}$, and $\mathbb{M}_n$ is replaced by the space $\mathcal{L}(\mathcal{H})$ of bounded operators in a Hilbert space $\mathcal{H}$. This theorem is the Hahn-Banach theorem of noncommutative analysis.

Theorem 3.1.1 is Stinespring's theorem restricted to algebras of matrices. It was proved by M.-D. Choi, *Completely positive linear maps* *on complex matrices,* Linear Algebra Appl., 10 (1975) 285–290, and by K. Kraus, *General state changes in quantum theory,* Ann. of Phys., 64 (1971) 311–335. It seems that the first paper has been well known to operator theorists and the second to physicists. The recent developments in quantum computation and quantum information theory have led to a renewed interest in these papers.

The book M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information,* Cambridge University Press, 2000, is a popular introduction to this topic. An older book from the physics literature is K. Kraus, *States, Effects, and Operations: Fundamental Notions of Quantum Theory,* Lecture Notes in Physics Vol. 190, Springer, 1983.

A positive matrix of trace one is called a *density matrix* in quantum mechanics. It is the noncommutative analogue of a probability distribution (a vector whose coordinates are nonnegative and add up to one). The requirement that density matrices are mapped to density matrices leads to the notion of a trace-preserving positive map. That this should happen also when the original system is tensored with another system (put in a larger system) leads to trace-preserving completely positive maps. Such maps are called *quantum channels.* Thus quantum channels are maps of the form (3.3) with the additional requirement $\sum V_j V_j^* = I$. The operators V_j are called the *noise operators,* or *errors* of the channel.

The representation (3.3) is one reason for the wide use of completely positive maps. Attempts to obtain some good representation theorem for positive maps were not very successful. See E. Størmer, *Positive linear maps of operator algebras,* Acta Math., 110 (1963) 233–278, S. L. Woronowicz, *Positive maps of low dimensional matrix algebras,* Reports Math. Phys., 10 (1976) 165–183, and M.-D. Choi, *Some assorted inequalities for positive linear maps on C^*-algebras,* J. Operator Theory, 4 (1980) 271–285. Let us say that a positive linear map $\Phi : \mathbb{M}_n \to \mathbb{M}_k$ is *decomposable* if it can be written as

$$\Phi(A) = \sum_{i=1}^{r} V_i^* A V_i + \sum_{j=1}^{s} W_j^* A^{\mathrm{t}} W_j.$$

If every positive linear map were decomposable it would follow that every real polynomial in n variables that takes only nonnegative values is a sum of squares of real polynomials. That the latter statement is false was shown by David Hilbert. The existence of a counterexample to the question on positive linear maps gives an easy proof of this result of Hilbert. See M.-D. Choi, *Positive linear maps,* cited in

Chapter 2, for a discussion.

The results of Exercises 3.2.2 and 3.2.3 are due to Choi and are given in his 1980 paper cited above. The idea that positive maps have a restricted 2-positive behavior seems to have first appeared in T. Ando, *Concavity of certain maps ...,* Linear Algebra Appl., 26 (1979) 203–241. Examples of maps on $\mathbb{M}_n$ that are $(n-1)$-positive but not n-positive were given in M.-D. Choi, *Positive linear maps on C^*-algebras,* Canadian J. Math., 24 (1972) 520–529. The simplest examples are of the type given in Exercise 3.2.4 (with n and $(n-1)$ in place of 3 and 2, respectively).

The Schwarz inequality is one of the most important and useful inequalities in mathematics. It is natural to seek its extensions in all directions and to expect that they will be useful. The reader should see the book J. M. Steele, *The Cauchy-Schwarz Master Class,* Math. Association of America, 2004, for various facets of the Schwarz inequality. (Noncommutative or matrix versions are not included.) Section IX.5 of MA is devoted to certain Schwarz inequalities for matrices. The operator inequality (3.19) was first proved for special types of positive maps (including completely positive ones) by E. H. Lieb and M. B. Ruskai, *Some operator inequalities of the Schwarz type,* Adv. Math., 12 (1974) 269–273. That 2-positivity is an adequate assumption was noted by Choi in his 1980 paper. Theorem 3.3.1 was proved in R. Bhatia and C. Davis, *More operator versions of the Schwarz inequality,* Commun. Math. Phys., 215 (2000) 239–244. It was noted there (observation due to a referee) that 4-positivity of Φ is adequate to ensure the validity of (3.16). That 3-positivity suffices but 2-positivity does not was observed by R. Mathias, *A note on: "More operator versions of the Schwarz inequality,"* Positivity, 8 (2004) 85–87. The inequalities (3.23) and (3.25) are proved in the paper of Bhatia and Davis cited above, and in a slightly different form in S.-G. Wang and W.-C. Ip, *A matrix version of the Wielandt inequality and its applications,* Linear Algebra Appl., 296 (1999) 171–181.

Section 3.4 is based on material in the paper V. I. Paulsen, S. C. Power, and R. R. Smith, *Schur products and matrix completions,* J. Funct. Anal., 85 (1989) 151–178, and on Paulsen's two books cited earlier. Theorem 3.4.3 is attributed to U. Haagerup, *Decomposition of completely bounded maps on operator algebras,* unpublished report. Calculating the exact value of the norm of a linear operator on a Hilbert space is generally a difficult problem. Calculating its norm as a Schur multiplier is even more difficult. Haagerup's Theorem gives some methods for such calculations.

Completion problems of various kinds have been studied by several authors with diverse motivations coming from operator theory, electrical engineering, and optimization. A helpful introduction may be obtained from C. R. Johnson, *Matrix completion problems: a survey,* Proc. Symposia in Applied Math. Vol. 40, American Math. Soc., 1990.

Theorem 3.5.1 was proved by T. Ando, *Structure of operators with numerical radius one,* Acta Sci. Math. (Szeged), 34 (1973) 11–15. The proof given here is different from the original one, and is from T. Ando, *Operator Theoretic Methods for Matrix Inequalities,* Sapporo, 1998. Theorem 3.5.3 is proved in T. Ando and K. Okubo, *Induced norms of the Schur multiplier operator,* Linear Algebra Appl., 147 (1991) 181–199. This and Haagerup's theorem are included in Ando's 1998 report from which we have freely borrowed. A lot more information about inequalities for Schur products may be obtained from this report.

The inequality (3.20) is called Berger's theorem. The lack of submultiplicativity and of its weaker substitutes has been a subject of much investigation in the theory of the numerical radius.

We have seen that even under the stringent assumption $AB = BA$ we need not have $w(AB) \leq w(A)w(B)$. Even the weaker assertion $w(AB) \leq \|A\|w(B)$ is not always true in this case. A 12×12 counterexample, in which $w(AB) > (1.01)\|A\|w(B)$ was found by V. Müller, *The numerical radius of a commuting product,* Michigan Math. J., 35 (1988) 255–260. This was soon followed by K. R. Davidson and J.A.R. Holbrook, *Numerical radii of zero-one matrices,* ibid., 35 (1988) 261–267, who gave a simpler 9×9 example in which $w(AB) > C\|A\|w(B)$ where $C = 1/\cos(\pi/9) > 1.064$. The reader will find in this paper a comprehensive discussion of the problem and its relation to other questions in dilation theory.

The formula (3.28) occurs in R. Bhatia, M.-D. Choi, and C. Davis, *Comparing a matrix to its off-diagonal part,* Oper. Theory: Adv. and Appl., 40 (1989) 151–164. The results of Exercises 3.6.2–3.6.6 are also taken from this paper. The ideas of this paper are taken further in R. Bhatia, *Pinching, trimming, truncating and averaging of matrices,* Am. Math. Monthly, 107 (2000) 602–608. Finding the exact norm of the operator $\triangle_n$ of Exercise 3.6.8 is hard. It is a well-known and important result of operator theory that for large n, the norm $\|\triangle_n\|$ is close to $\log n$. See the paper by R. Bhatia (2000) cited above.

The operation of replacing the matrix entries A_{ij} of a block matrix $[[A_{ij}]]$ by $f(A_{ij})$ for various functions f has been studied by several linear algebraists. See, for example, J. De Pillis, *Transformations on partitioned matrices,* Duke Math. J., 36 (1969) 511–515,

R. Merris, *Trace functions I,* ibid., 38 (1971) 527–530, and M. Marcus and W. Watkins, *Partitioned Hermitian matrices,* ibid., 38(1971) 237–249. Results of Exercises 3.6.11-3.6.13 are noted in this paper of Marcus and Watkins. Two foundational papers on this topic that develop a general theory are T. Ando and M.-D. Choi, *Non-linear completely positive maps,* in *Aspects of Positivity in Functional Analysis,* North-Holland Mathematical Studies Vol. 122, 1986, pp.3–13, and W. Arveson, *Nonlinear states on C^*-algebras,* in *Operator Algebras and Mathematical Physics,* Contemporary Mathematics Vol. 62, American Math. Society, 1987, pp. 283–343. Characterisations of nonlinear completely positive maps and Stinespring-type representation theorems are proved in these papers. These are substantial extensions of the representation (1.40). Exercise 3.6.14 is borrowed from the paper of Ando and Choi.

Finally, we mention that the theory of completely positive maps is now accompanied by the study of completely bounded maps, just as the study of positive measures is followed by that of bounded measures. The two books by Paulsen are an excellent introduction to the major themes of this subject. The books K. R. Parthasarathy, *An Introduction to Quantum Stochastic Calculus,* Birkhäuser, 1992, and P. A. Meyer, *Quantum Probability for Probabilists,* Lecture Notes in Mathematics Vol. 1538, Springer, 1993, are authoritative introductions to noncommutative probability, a subject in which completely positive maps play an important role.

Chapter Four

Matrix Means

Let a and b be positive numbers. Their arithmetic, geometric, and harmonic means are the familiar objects

$$A(a,b) = \frac{a+b}{2},$$
$$G(a,b) = \sqrt{ab},$$
$$H(a,b) = \left(\frac{a^{-1}+b^{-1}}{2}\right)^{-1}.$$

These have several properties that any object that is called a *mean* $M(a,b)$ should have. Some of these properties are

(i) $M(a,b) > 0$,

(ii) If $a \le b$, then $a \le M(a,b) \le b$,

(iii) $M(a,b) = M(b,a)$ (symmetry),

(iv) $M(a,b)$ is monotone increasing in a, b,

(v) $M(\alpha a, \alpha b) = \alpha M(a,b)$ for all positive numbers a, b, and α,

(vi) $M(a,b)$ is continuous in a, b.

The three of the most familiar means listed at the beginning satisfy these conditions. We have the inequality

$$H(a,b) \;\le\; G(a,b) \;\le\; A(a,b). \tag{4.1}$$

Among other means of a, b is the *logarithmic mean* defined as

$$L(a,b) \;=\; \frac{a-b}{\log a - \log b} \;=\; \int_0^1 a^t b^{1-t} dt. \tag{4.2}$$

This has the properties (i)–(vi) listed above. Further

$$G(a,b) \ \leq \ L(a,b) \ \leq \ A(a,b). \tag{4.3}$$

This is a refinement of the arithmetic-geometric mean inequality—the second part of (4.1). See Exercise 4.5.5 and Lemma 5.4.5.

Averaging operations are of interest in the context of matrices as well, and various notions of means of positive definite matrices A and B have been studied. A mean $M(A,B)$ should have properties akin to (i)–(vi) above. The order "$\leq$" now is the natural order $X \leq Y$ on Hermitian matrices. It is obvious what the analogues of properties (i)–(vi) are for the case of positive definite matrices. Property (v) has another interpretation: for positive numbers a, b and any nonzero complex number x

$$M(\bar{x}ax, \bar{x}bx) = \bar{x}M(a,b)x.$$

It is thus natural to expect any mean $M(A,B)$ to satisfy the condition

$$(\mathrm{v}') \qquad M(X^*AX,\ X^*BX) = X^*M(A,B)X,$$

for all $A, B > O$ and all nonsingular X. This condition is called *congruence invariance* and if the equality (v′) is true, we say that M is *invariant under congruence.* Restricting X to scalar matrices we see that

$$M(\alpha A, \alpha B) = \alpha M(A,B)$$

for all positive numbers α.

So we say that a *matrix mean* is a binary operation $(A,B) \longmapsto M(A,B)$ on the set of positive definite matrices that satisfies (the matrix versions of) the conditions (i)–(vi), the condition (v) being replaced by (v′).

What are good examples of such means? The arithmetic mean presents no difficulties. It is obvious that $M(A,B) = \frac{1}{2}(A+B)$ has all the six properties listed above. The harmonic mean of A and B should be the matrix $\left(\frac{A^{-1}+B^{-1}}{2}\right)^{-1}$. Now some of the properties (i)–(vi) are obvious, others are not. It is not clear what object should be called the geometric mean in this case. The product $A^{1/2}B^{1/2}$ is not Hermitian, let alone positive, unless A and B commute.

In this chapter we define a geometric mean of positive matrices and study its properties along with those of the arithmetic and the harmonic mean. We use these ideas to prove some theorems on operator monotonicity and convexity. These theorems are then used to derive important properties of the quantum entropy. A positive matrix in this chapter is assumed to be *strictly* positive. Extensions of some of the considerations to positive semidefinite matrices are briefly indicated.

4.1 THE HARMONIC MEAN AND THE GEOMETRIC MEAN

The *parallel sum* of two positive matrices A, B is defined as the matrix

$$A : B = (A^{-1} + B^{-1})^{-1}. \tag{4.4}$$

This definition could be extended to positive semidefinite matrices A, B by a limit from above:

$$A : B = \lim_{\varepsilon \downarrow 0} \left[(A + \varepsilon I)^{-1} + (B + \varepsilon I)^{-1}\right]^{-1} \text{ if } A, B \geq O. \tag{4.5}$$

This operation was studied by Anderson and Duffin in connection with electric networks. (If two wires with resistances r_1 and r_2 are connected in parallel, then their total resistance r according to one of Kirchhoff's laws is given by $\frac{1}{r} = \frac{1}{r_1} + \frac{1}{r_2}$.)

The *harmonic mean* of A, B is the matrix $2(A : B)$. To save on symbols we will not introduce a separate notation for it. Note that

$$\begin{aligned} A : B = (A^{-1} + B^{-1})^{-1} &= \left[A^{-1}(A + B)B^{-1}\right]^{-1} = B(A + B)^{-1}A \\ &= B(A + B)^{-1}A + B(A + B)^{-1}B - B(A + B)^{-1}B \\ &= B - B(A + B)^{-1}B. \end{aligned} \tag{4.6}$$

By symmetry

$$A : B = A - A(A + B)^{-1}A. \tag{4.7}$$

Thus $A : B$ is the *Schur complement* of $A + B$ in either of the block matrices

$$\begin{bmatrix} A & A \\ A & A + B \end{bmatrix} \text{ or } \begin{bmatrix} B & B \\ B & A + B \end{bmatrix}.$$

Several properties of $A : B$ can be derived from this.

4.1.1 Theorem

For any two positive matrices A, B we have

(i) $A : B \leq A$, $A : B \leq B$.

(ii) $A : B$ *is monotonically increasing and jointly concave in the arguments* A, B.

(iii)

$$A : B = \max\left\{Y : Y \geq O, \begin{bmatrix} A & A \\ A & A+B \end{bmatrix} \geq \begin{bmatrix} Y & O \\ O & O \end{bmatrix}\right\}. \quad (4.8)$$

Proof.

(i) The subtracted terms in (4.6) and (4.7) are positive.

(ii) See Corollary 1.5.3.

(iii) See Corollary 1.5.5. ■

4.1.2 Proposition

If $A \leq B$, then $A \leq 2(A : B) \leq B$.

Proof.

$$\begin{aligned} A \leq B &\Rightarrow 2A \leq A + B \\ &\Rightarrow 2(A+B)^{-1} \leq A^{-1} \\ &\Rightarrow 2A(A+B)^{-1}A \leq A \\ &\Rightarrow A = 2A - A \leq 2A - 2A(A+B)^{-1}A = 2(A : B). \end{aligned}$$

A similar argument shows $2(A : B) \leq B$. ■

Thus the harmonic mean satisfies properties (i)–(v) listed at the beginning of the chapter. (Notice one difference: for positive numbers a, b either $a \leq b$ or $b \leq a$; this is not true for positive matrices A, B.)

How about the geometric mean of A, B? If A, B commute, then their geometric mean can be defined as $A^{1/2}B^{1/2}$. But this is the trivial case. In all other cases this matrix is not even Hermitian. The matrix $\frac{1}{2}\left(A^{1/2}B^{1/2}+B^{1/2}A^{1/2}\right)$ is Hermitian but not always positive. Positivity is restored if we consider

$$\frac{1}{2}(B^{1/4}A^{1/2}B^{1/4}+A^{1/4}B^{1/2}A^{1/4}).$$

It turns out that this is not monotone in A, B. (Exercise: construct a 2×2 example to show this.) One might try other candidates; e.g., $e^{(\log A+\log B)/2}$, that reduce to $a^{1/2}b^{1/2}$ for positive numbers. This particular one is not monotone.

Here the property (v′)—congruence invariance—that we expect a mean to have is helpful. We noted in Exercise 1.6.1 that any two positive matrices are simultaneously congruent to diagonal matrices. The geometric mean of two positive diagonal matrices A and B, naturally, is $A^{1/2}B^{1/2}$.

Let us introduce a notation and state a few elementary facts that will be helpful in the ensuing discussion. Let $GL(n)$ be the group consisting of $n\times n$ invertible matrices. Each element X of $GL(n)$ gives a congruence transformation on $\mathbb{M}_n$. We write this as

$$\Gamma_X(A)=X^*AX. \tag{4.9}$$

The collection $\{\Gamma_X : X\in GL(n)\}$ is a group of transformations on $\mathbb{M}_n$. We have $\Gamma_X\Gamma_Y=\Gamma_{YX}$ and $\Gamma_X^{-1}=\Gamma_{X^{-1}}$. This group preserves the set of positive matrices. Given a pair of matrices A, B we write $\Gamma_X(A,B)$ for $(\Gamma_X(A),\Gamma_X(B))$.

Let A, B be positive matrices. Then

$$\Gamma_{A^{-1/2}}(A,B)=(I,A^{-1/2}BA^{-1/2}).$$

We can find a unitary matrix U such that $U^*\left(A^{-1/2}BA^{-1/2}\right)U=D$, a diagonal matrix. So

$$\Gamma_{A^{-1/2}U}(A,B)=(I,D).$$

The geometric mean of the matrices I and D is

$$D^{1/2}=U^*\left(A^{-1/2}BA^{-1/2}\right)^{1/2}U.$$

So, if the geometric mean of two positive matrices A and B is required to satisfy the property (v′), then it has to be the matrix

$$A\#B=A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^{1/2}A^{1/2}. \tag{4.10}$$

If A and B commute, then $A\#B = A^{1/2}B^{1/2}$. The expression (4.10) does not appear to be symmetric in A and B. However, it is. This is seen readily from another description of $A\#B$. By Exercise 1.2.13 the matrix in (4.10) is the unique positive solution of the equation

$$XA^{-1}X = B. \tag{4.11}$$

If we take inverses of both sides, then this equation is transformed to $XB^{-1}X = A$. This shows that

$$A\#B = B\#A. \tag{4.12}$$

Using Theorem 1.3.3 and the relation

$$A = (A\#B)B^{-1}(A\#B)$$

that we have just proved, we see that

$$\begin{bmatrix} A & A\#B \\ A\#B & B \end{bmatrix} \geq O. \tag{4.13}$$

On the other hand if X is any Hermitian matrix such that

$$\begin{bmatrix} A & X \\ X & B \end{bmatrix} \geq O, \tag{4.14}$$

then again by Theorem 1.3.3, we have $A \geq XB^{-1}X$. Hence

$$B^{-1/2}AB^{-1/2} \geq B^{-1/2}XB^{-1}XB^{-1/2} = (B^{-1/2}XB^{-1/2})^2.$$

Taking square roots and then applying the congruence $\Gamma_{B^{1/2}}$, we get from this

$$B^{1/2}(B^{-1/2}AB^{-1/2})^{1/2}B^{1/2} \geq X.$$

In other words $A\#B \geq X$ for any Hermitian matrix X that satisfies the inequality (4.14).

The following theorem is a summary of our discussion so far.

4.1.3 Theorem

Let A and B be two positive matrices. Let

$$A\#B = A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2}.$$

Then

(i) $A\#B = B\#A$,

(ii) *$A\#B$ is the unique positive solution of the equation $XA^{-1}X = B$,*

(iii) *$A\#B$ has an extremal property:*

$$A\#B = \max\left\{X : X = X^*, \begin{bmatrix} A & X \\ X & B \end{bmatrix} \geq O\right\}. \tag{4.15}$$

The properties (i)–(vi) listed at the beginning of the chapter can be verified for $A\#B$ using one of the three characterizations given in Theorem 4.1.3. Thus, for example, the symmetry (4.12) is apparent from (4.15) as well. Monotonicity in the variable B is apparent from (4.10) and Proposition 1.2.9; and then by symmetry we have monotonicity in A. This is plain from (4.15) too. From (4.15) we see that $A\#B$ is jointly concave in A and B.

Since congruence operations preserve order, the inequality (4.1) is readily carried over to operators. We have

$$2(A : B) \leq A\#B \leq \frac{1}{2}(A + B). \tag{4.16}$$

It is easy to see either from (4.10) or from Theorem 4.1.3 (ii) that

$$A^{-1}\#B^{-1} = (A\#B)^{-1}. \tag{4.17}$$

4.1.4 Exercise

Use the characterization (4.15) and the symmetry (4.12) to give another proof of the second inequality in (4.16). Use (4.15) and (4.17) to give another proof of the first inequality in (4.16).

If A or B is not strictly positive, we can define their geometric mean by a limiting procedure, as we did in (4.5) for the parallel sum.

The next theorem describes the effect of positive linear maps on these means.

4.1.5 Theorem

Let Φ be any positive linear map on $\mathbb{M}_n$. Then for all positive matrices A, B

(i) $\Phi(A : B) \;\leq\; \Phi(A) : \Phi(B)$;

(ii) $\Phi(A\#B) \;\leq\; \Phi(A)\#\Phi(B)$.

Proof. (i) By the extremal characterization (4.8)

$$\begin{bmatrix} A - (A:B) & A \\ A & A+B \end{bmatrix} \geq O.$$

By Exercise 3.2.2 (ii), we get from this

$$\begin{bmatrix} \Phi(A) - \Phi(A:B) & \Phi(A) \\ \Phi(A) & \Phi(A)+\Phi(B) \end{bmatrix} \geq O.$$

Again, by (4.8) this means $\Phi(A:B) \leq \Phi(A):\Phi(B)$.

The proof of (ii) is similar to this. Use the extremal characterization (4.15) for $A\#B$, and Exercise 3.2.2 (ii). ■

For the special map $\Gamma_X(A) = X^*AX$, where X is any invertible matrix the two sides of (i) and (ii) in Theorem 4.1.5 are equal.This need not be the case if X is not invertible.

4.1.6 Exercise

Let A, B, and X be the 2×2 matrices

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 20 & 6 \\ 6 & 2 \end{bmatrix}, \quad X = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Show that

$$X^*(A\#B)X = \begin{bmatrix} 8 & 0 \\ 0 & 0 \end{bmatrix}, \quad (X^*AX)\#(X^*BX) = \begin{bmatrix} \sqrt{80} & 0 \\ 0 & 0 \end{bmatrix}.$$

So, if $\Phi(A) = X^*AX$, then in this example we have

$$\Phi(A\#B) \neq \Phi(A)\#\Phi(B).$$

The inequality (4.13) and Proposition 1.3.2 imply that there exists a contraction K such that $A\#B = A^{1/2}KB^{1/2}$. More is true as the next Exercise and Proposition show.

4.1.7 Exercise

Let $U = (A^{-1/2}BA^{-1/2})^{1/2}A^{1/2}B^{-1/2}$. Show that $U^*U = UU^* = I$. Thus we can write

$$A\#B = A^{1/2}UB^{1/2},$$

where U is unitary.

It is an interesting fact that this property characterizes the geometric mean:

4.1.8 Proposition

Let A, B be positive matrices and suppose U is a unitary matrix such that $A^{1/2}UB^{1/2}$ is positive. Then $A^{1/2}UB^{1/2} = A\#B$.

Proof. Let $G = A^{1/2}UB^{1/2}$. Then

$$\begin{bmatrix} A & G \\ G & B \end{bmatrix} = \begin{bmatrix} A^{1/2} & O \\ O & B^{1/2} \end{bmatrix} \begin{bmatrix} I & U \\ U^* & I \end{bmatrix} \begin{bmatrix} A^{1/2} & O \\ O & B^{1/2} \end{bmatrix} \sim \begin{bmatrix} I & U \\ U^* & I \end{bmatrix}.$$

We have another congruence

$$\begin{bmatrix} A & G \\ G & B \end{bmatrix} \sim \begin{bmatrix} A - GB^{-1}G & O \\ O & B \end{bmatrix}.$$

(See the proof of Theorem 1.3.3.) Note that the matrix $\begin{bmatrix} I & U \\ U^* & I \end{bmatrix}$ has rank n. Since congruence preserves rank we must have $A = GB^{-1}G$. But then, by Theorem 4.1.3 (ii), $G = A\#B$. ■

Two more ways of expressing the geometric mean are given in the following propositions. We use here the fact that if X is a matrix with positive eigenvalues, then it has a unique square root Y with positive eigenvalues. A proof is given in Exercise 4.5.2.

4.1.9 Proposition

Let A, B be positive matrices and let $\left(A^{-1}B\right)^{1/2}$ be the the square root of $A^{-1}B$ that has positive eigenvalues. Then

$$A\#B = A\left(A^{-1}B\right)^{1/2}.$$

Proof. We have the identity

$$A^{-1/2}BA^{-1/2} = A^{1/2}A^{-1}BA^{-1/2} = \left[A^{1/2}\left(A^{-1}B\right)^{1/2}A^{-1/2}\right]^2.$$

Taking square roots, we get

$$\left(A^{-1/2}BA^{-1/2}\right)^{1/2} = A^{1/2}\left(A^{-1}B\right)^{1/2}A^{-1/2}.$$

This, in turn, shows that

$$A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^{1/2}A^{1/2}=A\left(A^{-1}B\right)^{1/2}. \qquad \blacksquare$$

4.1.10 Exercise

Show that for positive matrices A, B we have

$$A\#B=\left(AB^{-1}\right)^{1/2}B.$$

4.1.11 Proposition

Let A, B be positive matrices. Then

$$A\#B=(A+B)\left[(A+B)^{-1}A(A+B)^{-1}B\right]^{1/2}.$$

(The matrix inside the square brackets has positive eigenvalues and the square root chosen is the one with positive eigenvalues.)

Proof. Use the identity

$$X=\left(X^{-1}+I\right)^{-1}(I+X)$$

to get

$$\begin{aligned}A^{-1}B&=\left(B^{-1}A+I\right)^{-1}\left(I+A^{-1}B\right)\\&=(A+B)^{-1}\left(AB^{-1}\right)^{-1}(A+B).\end{aligned}$$

Taking square roots, we get

$$\left(A^{-1}B\right)^{1/2}=(A+B)^{-1}\left(AB^{-1}\right)^{-1/2}(A+B).$$

This gives

$$A\left(A^{-1}B\right)^{1/2}(A+B)^{-1}\left(AB^{-1}\right)^{1/2}B=A(A+B)^{-1}B.$$

Using Proposition 4.1.9 and Exercise 4.1.10, we get from this

$$(A\#B)(A+B)^{-1}(A\#B)=A(A+B)^{-1}B.$$

Premultiply both sides by $(A+B)^{-1}$, and then take square roots, to get

$$(A+B)^{-1}(A\#B)=\left[(A+B)^{-1}A(A+B)^{-1}B\right]^{1/2}.$$

This proves the proposition. ■

On first sight, the three expressions in 4.1.9–4.1.11 do not seem to be positive matrices, nor do they seem to be symmetric in A, B.

The expression (4.10) and the ones given in Propositions 4.1.9 and 4.1.11 involve finding square roots of matrices, as should be expected in any definition of geometric mean. Calculating these square roots is not an easy task. For 2×2 matrices we have a formula that makes computation easier.

4.1.12 Proposition

Let A and B be 2×2 positive matrices each of which has determinant one. Then

$$A\#B = \frac{A+B}{\sqrt{\det(A+B)}}.$$

Proof. Use the formula given for $A\#B$ in Proposition 4.1.9. Let $X = (A^{-1}B)^{1/2}$. Then $\det X = 1$ and so X has two positive eigenvalues λ and $1/\lambda$. Further,

$$\det(A+B) = \det\left[A(I + A^{-1}B)\right] = \det(I + X^2) = (\lambda + 1/\lambda)^2,$$

and hence $\operatorname{tr} X = \sqrt{\det(A+B)}$. So, by the Cayley-Hamilton theorem

$$X^2 - \sqrt{\det(A+B)}X + I = O.$$

Multiply on the left by A and rearrange terms to get

$$A(A^{-1}B)^{1/2} = \frac{A+B}{\sqrt{\det(A+B)}}. \quad ■$$

Exercise. Let A and B be 2×2 positive matrices and let $\det A = \alpha^2$, $\det B = \beta^2$. Then

$$A\#B = \frac{\sqrt{\alpha\beta}}{\sqrt{\det(\alpha^{-1}A + \beta^{-1}B)}}(\alpha^{-1}A + \beta^{-1}B).$$

4.2 SOME MONOTONICITY AND CONVEXITY THEOREMS

In Section 5 of Chapter 1 and Section 7 of Chapter 2 we have discussed the notions of monotonicity, convexity and concavity of operator functions. Operator means give additional information as well as more insight into these notions. Some of the theorems in this section have been proved by different arguments in Chapter 1.

4.2.1 Theorem

If $A \geq B \geq O$, then $A^r \geq B^r$ for all $0 \leq r \leq 1$.

Proof. We know that the assertion is true for $r = 0, 1$. Suppose r_1, r_2 are two real numbers for which $A^{r_1} \geq B^{r_1}$ and $A^{r_2} \geq B^{r_2}$. Then, by monotonicity of the geometric mean, we have $A^{r_1} \# A^{r_2} \geq B^{r_1} \# B^{r_2}$. This is the same as saying $A^{(r_1+r_2)/2} \geq B^{(r_1+r_2)/2}$. Thus, the set of real numbers r for which $A^r \geq B^r$ is a closed convex set. Since 0 and 1 belong to this set, so does the entire interval $[0, 1]$. ■

4.2.2 Exercise

We know that the function $f(t) = t^2$ is not matrix monotone of order 2. Show that the function $f(t) = t^r$ on $\mathbb{R}_+$ is not matrix monotone of order 2 for any $r > 1$. [Hint: Prove this first for $r > 2$.]

It is known that a function f from $\mathbb{R}_+$ into itself is operator monotone if and only if it is operator concave. For the functions $f(t) = t^r$, $0 \leq r \leq 1$, operator concavity is easily proved:

4.2.3 Theorem

For $0 < r < 1$, the map $A \longmapsto A^r$ on positive matrices is concave.

Proof. Use the representation

$$A^r = \int_0^\infty A(\lambda + A)^{-1}\, d\mu(\lambda), \qquad 0 < r < 1 \tag{4.18}$$

(see Theorem 1.5.8). The integrand

$$A(\lambda + A)^{-1} = (\lambda A^{-1} + I)^{-1} = \frac{A}{\lambda} : I$$

is concave in A by Theorem 4.1.1. Hence, so is the integral. (The integrand is also monotone in A. In Section 1.5 we used this argument to prove Theorem 4.2.1 and some other statements.) ■

4.2.4 Exercise

For $0 < r < 1$, the map $A \longmapsto A^{-r}$ on positive matrices is monotone decreasing and convex. [Use the facts that $A \longmapsto A^r$ is monotone and concave, and $A \longmapsto A^{-1}$ is monotone decreasing and convex.] See Exercise 1.5.10 also.

4.2.5 Exercise

The map $A \longmapsto \log A$ on positive matrices is monotone and concave.

[Hint: $\frac{d}{dr}\Big|_{r=0^+} a^r = \log a$.]

4.2.6 Exercise

The map $A \longmapsto -A \log A$ on positive matrices is concave.

[Hint: $\frac{d}{dr}\Big|_{r=1^+} a^r = a \log a$. Use Theorem 1.5.8.]

Some results on convexity of tensor product maps can be deduced easily using the harmonic mean. The following theorem was proved by Lieb. This formulation and proof are due to Ando.

4.2.7 Theorem

For $0 < r < 1$, the map $f(A, B) = A^r \otimes B^{1-r}$ is jointly concave and monotone on pairs of positive matrices A, B.

Proof. Note that $A^r \otimes B^{1-r} = (I \otimes B)(A \otimes B^{-1})^r$. So, by the representation (4.18) we have

$$f(A, B) = \int_0^\infty (I \otimes B)(A \otimes B^{-1})(\lambda I \otimes I + A \otimes B^{-1})^{-1} \, d\mu(\lambda).$$

The integrand can be written as

$$(\lambda A^{-1} \otimes I + I \otimes B^{-1})^{-1} \;=\; \left(\frac{A \otimes I}{\lambda}\right) : (I \otimes B).$$

This is monotone and jointly concave in A, B. Hence, so is $f(A, B)$. ∎

4.2.8 Exercise

Let r_1, r_2 be positive numbers with $r_1 + r_2 \leq 1$. Show that $f(A, B) = A^{r_1} \otimes B^{r_2}$ is jointly concave and monotone on pairs of positive matrices A, B. [Hint: Let $r_1 + r_2 = r$. Then $f(A, B) \;=\; (A^{r_1/r} \otimes B^{r_2/r})^r$.]

4.2.9 Exercise

Let $r_1, r_2, \ldots, r_k$ be positive numbers with $r_1 + \cdots + r_k = 1$. Then the product $A_1^{r_1} \otimes A_2^{r_2} \otimes \cdots \otimes A_k^{r_k}$ is jointly concave on k-tuples of positive matrices $A_1, \ldots, A_k$.

A special case of this says that for $k = 1, 2, \ldots,$ the map $A \longmapsto \otimes^k A^{1/k}$ on positive matrices is concave. This leads to the inequality

$$\otimes^k (A + B)^{1/k} \;\geq\; \otimes^k A^{1/k} + \otimes^k B^{1/k}. \tag{4.19}$$

By restricting to symmetry classes of tensors one obtains inequalities for other induced operators. For example

$$\wedge^k (A + B)^{1/k} \;\geq\; \wedge^k A^{1/k} + \wedge^k B^{1/k}. \tag{4.20}$$

For $k = n$, this reduces to the famous *Minkowski determinant inequality*

$$\det(A + B)^{1/n} \;\geq\; \det A^{1/n} + \det B^{1/n}. \tag{4.21}$$

It is clear that many inequalities of this kind are included in the master inequality (4.19).

4.2.10 Exercise

For $0 \leq r \leq 1$, the map $f(A, B) = A^{-r} \otimes B^{1+r}$ is jointly convex on pairs of positive matrices A, B. [Hint: $f(A, B) \;=\; \int_0^\infty (I \otimes B)(\lambda A \otimes I + I \otimes B)^{-1}(I \otimes B)\, d\mu(\lambda)$.]

4.3 SOME INEQUALITIES FOR QUANTUM ENTROPY

Theorem 4.2.7 is equivalent to a theorem of Lieb on the concavity of one of the matrix functions arising in the study of entropy in quantum mechanics. We explain this and related results briefly.

Let $(p_1, \ldots, p_k)$ be a probability vector; i.e., $p_j \geq 0$ and $\sum p_j = 1$. The function

$$H(p_1, \ldots, p_k) \; = \; -\sum p_j \log p_j, \tag{4.22}$$

called the *entropy function*, is of fundamental importance in information theory.

In quantum mechanics, the role analogous to that of $(p_1, \ldots, p_k)$ is played by a positive matrix A with $\operatorname{tr} A = 1$. Such a matrix is called a *density matrix*. The *entropy* of A is defined as

$$S(A) \; = \; -\operatorname{tr}(A \log A). \tag{4.23}$$

The condition $\operatorname{tr} A = 1$ is not essential for *some* of our theorems; and we will drop it some times.

It is easy to see that the function (4.22) is jointly concave in p_j. An analogous result is true for the quantum mechanical entropy (4.23). In fact we have proved a much stronger result in Exercise 4.2.6. The proof of concavity of the scalar function (4.23) does not require the machinery of operator concave function.

4.3.1 Exercise

Let A be any Hermitian matrix and f any convex function on $\mathbb{R}$. Then for every unit vector x

$$f(\langle x, Ax\rangle) \; \leq \; \langle x, f(A)x\rangle. \tag{4.24}$$

Use this to prove that $S(A)$ is a concave function of A. [Hint: Choose an orthonormal basis $\{u_i\}$ consisting of eigenvectors of $\frac{1}{2}(A+B)$. Show that for any convex function f

$$\sum \left\langle u_i, f\left(\frac{A+B}{2}\right) u_i \right\rangle \leq \sum \left\langle u_i, \frac{f(A)+f(B)}{2} u_i \right\rangle .]$$

4.3.2 Proposition

Let f be a convex differentiable function on $\mathbb{R}$ and let A, B be any two Hermitian matrices. Then

$$\operatorname{tr}\left[f(A) - f(B)\right] \geq \operatorname{tr}\left[(A-B)f'(B)\right]. \tag{4.25}$$

Proof. Let $\{u_i\}$ be an orthonormal basis whose elements are eigenvectors of B corresponding to eigenvalues β_i. Then

$$\begin{aligned}\operatorname{tr}[f(A) - f(B)] &= \sum \langle u_i, [f(A) - f(B)]u_i\rangle \\ &= \sum [\langle u_i, f(A)u_i\rangle - f(\beta_i)] \\ &\geq \sum [f(\langle u_i, Au_i\rangle) - f(\beta_i)] \\ &\geq \sum [\langle u_i, Au_i\rangle - \beta_i] f'(\beta_i) \\ &= \sum \langle u_i, (A - B)f'(B)u_i\rangle \\ &= \operatorname{tr}[(A - B)f'(B)].\end{aligned}$$

The first inequality in this chain follows from (4.24) and the second from the convexity of f. ■

Other notions of entropy have been introduced in classical information theory and in quantum mechanics. One of them is called *skew entropy* or *entropy of A relative to K*, where A is positive and K Hermitian. This is defined as

$$S(A, K) = \frac{1}{2} \operatorname{tr} \left[A^{1/2}, K\right]^2. \tag{4.26}$$

More generally we may consider for $0 < t < 1$ the function

$$S_t(A, K) = \frac{1}{2} \operatorname{tr} \left[A^t, K\right] \left[A^{1-t}, K\right]. \tag{4.27}$$

Here $[X, Y]$ is the Lie bracket $XY - YX$. Note that

$$S_t(A, K) = \operatorname{tr} (KA^t\, KA^{1-t} - K^2 A). \tag{4.28}$$

The quantity (4.26) was introduced by Wigner and Yanase, (4.27) by Dyson. These are measures of the amount of noncommutativity of A with a fixed Hermitian operator K. The Wigner-Yanase-Dyson conjecture said $S_t(A, K)$ is concave in A for each K. Note that $\operatorname{tr}(-K^2 A)$

is linear in A. So, from the expression (4.28) we see that the conjecture says that for each K, and for $0 < t < 1$, the function $\operatorname{tr} KA^tKA^{1-t}$ is concave in A. A more general result was proved by Lieb in 1973.

4.3.3 Theorem (Lieb)

For any matrix X, and for $0 \le t \le 1$, the function

$$f(A,B) \;=\; \operatorname{tr} X^*A^tXB^{1-t}$$

is jointly concave on pairs of positive matrices A, B.

Proof. We will show that Theorems 4.2.7 and 4.3.3 are equivalent.

The tensor product $\mathcal{H} \otimes \mathcal{H}^*$ and the space $\mathcal{L}(\mathcal{H})$ are isomorphic. The isomorphism φ acts as $\varphi(x \otimes y^*) = xy^*$. If $e_i, 1 \le i \le n$ is the standard basis for $\mathcal{H}$, and E_{ij} the matrix units in $\mathcal{L}(\mathcal{H})$, then

$$\varphi(e_i \otimes e_j^*) = e_ie_j^* \;=\; E_{ij},$$

$$\varphi^{-1}(A) = \sum_j \,(Ae_j) \;\otimes\; e_j^*.$$

From this one can see that

$$\left\langle \varphi^{-1}(A_1), \varphi^{-1}(A_2) \right\rangle_{\mathcal{H}\otimes\mathcal{H}^*} \;=\; \operatorname{tr} A_1^*A_2 \;=\; \langle A_1, A_2 \rangle_{\mathcal{L}(\mathcal{H})}\,.$$

Thus φ is an isometric isomorphism between the Hilbert space $\mathcal{H} \otimes \mathcal{H}^*$ and the Hilbert space $\mathcal{L}(\mathcal{H})$ (the latter with the inner product $\langle A_1, A_2 \rangle_{\mathcal{L}(\mathcal{H})} = \operatorname{tr} A_1^*A_2$).

Let $\overline{\varphi}$ be the map induced by φ on operators; i.e., $\overline{\varphi}$ makes the diagram

$$\begin{array}{ccc}
\mathcal{H} \otimes \mathcal{H}^* & \xrightarrow{\quad\varphi\quad} & \mathcal{L}(\mathcal{H}) \\
\Big\downarrow A \otimes B^* & & \Big\downarrow \overline{\varphi}(A \otimes B^*) \\
\mathcal{H} \otimes \mathcal{H}^* & \xrightarrow[\quad\varphi\quad]{} & \mathcal{L}(\mathcal{H})
\end{array}$$

commute. It is easy to see that

$$\overline{\varphi}(A \otimes B^*)(T) \;=\; ATB^* \;\text{ for all } T \in \mathcal{L}(\mathcal{H}).$$

So

$$\begin{aligned}\operatorname{tr} X^*AXB^* &= \langle X, AXB^* \rangle_{\mathcal{L}(\mathcal{H})} \\ &= \langle X, \overline{\varphi}(A \otimes B^*)(X) \rangle_{\mathcal{L}(\mathcal{H})} \\ &= \left\langle \varphi^{-1}(X)\ ,\ (A \otimes B^*)\varphi^{-1}(X) \right\rangle_{\mathcal{H}\otimes\mathcal{H}^*}.\end{aligned}$$

Thus for positive A, B, the concavity of $\operatorname{tr} X^*A^tXB^{1-t}$ is equivalent to the concavity of $A^t \otimes B^{1-t}$. ■

Other useful theorems can be derived from Theorem 4.3.3. Here is an example. The concept of *relative entropy* in classical information theory is defined as follows. Let p, q be two probability distributions; i.e., $p = (p_1, \ldots, p_k)$, $q = (q_1, \ldots, q_k)$, $p_j \geq 0$, $q_j \geq 0$, $\sum p_j = \sum q_j = 1$. Their relative entropy is defined as

$$S(p|q) = \begin{cases} \sum_j p_j(\log p_j - \log q_j) & \text{if } p_j = 0 \text{ whenever } q_j = 0 \\ \\ +\infty, \ \text{ otherwise.} \end{cases}$$

The *relative entropy* of density matrices A, B is defined as

$$S(A|B) \;=\; \operatorname{tr} A(\log A - \log B). \tag{4.29}$$

4.3.4 Exercise (Klein's Inequality)

For positive matrices A, B

$$\operatorname{tr} A(\log A - \log B) \;\geq\; \operatorname{tr}(A - B). \tag{4.30}$$

[Hint: Use Proposition 4.3.2. with $f(x) = x \log x$]. Thus, if A, B are density matrices, then

$$S(A|B) \;\geq\; 0. \tag{4.31}$$

Note

$$S(A|I) \;=\; -S(A) \;=\; \operatorname{tr} A \log A. \tag{4.32}$$

(I is not a density matrix.) We have seen that $S(A|I)$ is a convex function of A.

4.3.5 Theorem

The relative entropy $S(A|B)$ is a jointly convex function of A and B.

Proof. For A positive and X arbitrary, let

$$\begin{aligned} I_t(A,X) &= \operatorname{tr}(X^*A^tXA^{1-t} - X^*XA), \\ I(A,X) &= \frac{d}{dt}\bigg|_{t=0^+} I_t(A,X). \end{aligned}$$

By Lieb's theorem $I_t(A,X)$ is a concave function of A. Hence, so is $I(A,X)$. Note that

$$I(A,X) \;=\; \operatorname{tr}(X^*(\log A)XA - X^*X(\log A)A).$$

Now, given the positive matrices A, B let T, X be the 2×2 block matrices

$$T \;=\; \begin{pmatrix} A & O \\ O & B \end{pmatrix}, \; X \;=\; \begin{pmatrix} O & O \\ I & O \end{pmatrix}.$$

Then $I(T,X) = -S(A|B)$. Since $I(T,X)$ is concave, $S(A|B)$ is convex. ■

The next few results describe the behaviour of the entropy function (4.23) with respect to tensor products. Here the condition $\operatorname{tr} A = 1$ will be crucial for some of the statements.

4.3.6 Proposition

Let A, B be any two density matrices. Then

$$S(A \otimes B) \;=\; S(A) + S(B). \tag{4.33}$$

Proof. Let A, B have eigenvalues $\lambda_1, \dots, \lambda_n$ and $\mu_1, \dots, \mu_m$, respectively. Then $\sum \lambda_i = 1$, and $\sum \mu_j = 1$. The tensor product $A \otimes B$ has eigenvalues $\lambda_i \mu_j$, $1 \le i \le n$, $1 \le j \le m$. So

$$\begin{aligned} S(A \otimes B) &= -\sum_{i,j} \lambda_i \mu_j \log(\lambda_i \mu_j) \\ &= -\sum_i \lambda_i \log \lambda_i - \sum \mu_j \log \mu_j \\ &= S(A) + S(B). \quad \blacksquare \end{aligned}$$

The equation (4.33) says that the entropy function S is *additive* with respect to tensor products.

Now let $\mathcal{H}_1, \mathcal{H}_2$ be two Hilbert spaces, and let $A \in \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_2)$. The operator A is called *decomposable* if $A = A_1 \otimes A_2$ where A_1, A_2 are operators on $\mathcal{H}_1, \mathcal{H}_2$. Not every operator on $\mathcal{H}_1 \otimes \mathcal{H}_2$ is decomposable. We associate with every operator A on $\mathcal{H}_1 \otimes \mathcal{H}_2$ two operators A_1, A_2 on $\mathcal{H}_1, \mathcal{H}_2$ called the *partial traces* of A. These are defined as follows.

Let $\dim \mathcal{H}_1 = n$, $\dim \mathcal{H}_2 = m$ and let $e_1, \dots, e_m$ be an orthonormal basis in $\mathcal{H}_2$. For $A \in \mathcal{L}(\mathcal{H}_1 \otimes \mathcal{H}_2)$ its partial trace $A_1 = \operatorname{tr}_{\mathcal{H}_2} A$ is an operator on $\mathcal{H}_1$ defined by the relation

$$\langle x, A_1 y \rangle = \sum_{i=1}^{m} \langle x \otimes e_i, A(y \otimes e_i) \rangle \tag{4.34}$$

for all $x, y \in \mathcal{H}_1$.

4.3.7 Exercise

The operator A_1 above is well defined. (It is independent of the basis $\{e_i\}$ chosen for $\mathcal{H}_2$.)

The partial trace $A_2 = \operatorname{tr}_{\mathcal{H}_1} A$ is defined in an analogous way.

It is clear that if A is positive, then so are the partial traces A_1, A_2; and if A is a density matrix, then so are A_1, A_2. Further, if $A = A_1 \otimes A_2$ and A_1, A_2 are density matrices, then $A_1 = \operatorname{tr}_{\mathcal{H}_2} A$, $A_2 = \operatorname{tr}_{\mathcal{H}_1} A$.

4.3.8 Exercise

Let A be an operator on $\mathcal{H}_1 \otimes \mathcal{H}_2$ with partial traces A_1, A_2. Then for every decomposable operator B of the form $B_1 \otimes I$ on $\mathcal{H}_1 \otimes \mathcal{H}_2$ we have

$$\operatorname{tr} AB \;=\; \operatorname{tr} A_1 B_1. \tag{4.35}$$

The next proposition is called the *subadditivity* property of entropy.

4.3.9 Proposition

Let A be a density matrix in $\mathcal{H}_1 \otimes \mathcal{H}_2$ with partial traces A_1, A_2. Then

$$S(A) \;\leq\; S(A_1) + S(A_2) = S(A_1 \otimes A_2). \tag{4.36}$$

Proof. The matrix $A_1 \otimes A_2$ is a density matrix. So, by Exercise 4.3.4, the relative entropy $S(A|A_1 \otimes A_2)$ is positive. By definition

$$\begin{aligned} S(A|A_1 \otimes A_2) &= \operatorname{tr} A(\log A - \log(A_1 \otimes A_2)) \\ &= \operatorname{tr} A(\log A - \log(A_1 \otimes I) - \log(I \otimes A_2)). \end{aligned}$$

By Exercise 4.3.8, this shows

$$\begin{aligned} S(A|A_1 \otimes A_2) &= \operatorname{tr} A \log A - \operatorname{tr} A_1 \log A_1 - \operatorname{tr} A_2 \log A_2 \\ &= -S(A) + S(A_1) + S(A_2). \end{aligned} \tag{4.37}$$

Since this quantity is positive, we have the inequality in (4.36). ■

There is another way of looking at the partial trace operation that is more transparent and makes several calculations easier:

4.3.10 Proposition

Let $f_1, \ldots, f_n$ and $e_1, \ldots, e_m$ be orthonormal bases for $\mathcal{H}_1$ and $\mathcal{H}_2$. Let A be an operator on $\mathcal{H}_1 \otimes \mathcal{H}_2$ and write its matrix in the basis $f_i \otimes e_j$ in the $n \times n$ partitioned form

$$A \;=\; [[A_{ij}]] \tag{4.38}$$

where $A_{ij}, 1 \leq i, j \leq n$ are $m \times m$ matrices. Then $\operatorname{tr}_{\mathcal{H}_2} A$ is the $n \times n$ matrix defined as

$$\text{tr}_{\mathcal{H}_2} A = [[\text{tr} A_{ij}]] . \quad (4.39)$$

Proof. It is not difficult to derive this relation from (4.34). ■

4.3.11 Exercise

The map $\text{tr}_{\mathcal{H}_2}$ is the composition of three special kinds of maps described below.

(i) Let $\omega = e^{2\pi i/m}$ and let $U = \text{diag}(1, \omega, \ldots, \omega^{m-1})$. Let W be the $n \times n$ block-diagonal matrix $W = U \oplus U \oplus \cdots \oplus U$ (n copies). Let

$$\Phi_1(A) = \frac{1}{m} \sum_{k=0}^{m-1} W^{*k} A W^k, \quad (4.40)$$

where A is as in (4.38). Show that

$$\Phi_1(A) = [[\text{diag}\ (A_{ij})]]. \quad (4.41)$$

(See (3.28)).

(ii) Let V be the $m \times m$ permutation matrix defined as

$$V = \begin{bmatrix} 0 & 0 & 0 & \ldots & 1 \\ 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot\cdot 1 & 0 \end{bmatrix}.$$

Let X be the $n \times n$ block-diagonal matrix $X = V \oplus V \oplus \cdots \oplus V$ (n *copies*). Let

$$\Phi_2(A) = \frac{1}{m} \sum_{k=0}^{m-1} X^{*k} A X^k. \quad (4.42)$$

Show that

$$\Phi_2(\Phi_1(A)) = \left[\left[\left(\frac{1}{m}\mathrm{tr}A_{ij}\right) I_m\right]\right]. \tag{4.43}$$

Thus the effect of Φ_2 on the block matrix (4.41) is to replace each of the diagonal matrices A_{ij} by the scalar matrix with the same trace as A_{ij}.

(iii) Let A be as in (4.38) and let $A_{ij}^{(1,1)}$ be the (1,1) entry of A_{ij}. Let

$$\Phi_3(A) = m\left[\left[A_{ij}^{(1,1)}\right]\right]. \tag{4.44}$$

Note that the matrix $\left[\left[A_{ij}^{(1,1)}\right]\right]$ is a principal $n \times n$ submatrix of A. We have then

$$\mathrm{tr}_{\mathcal{H}_2} A = \Phi_3\Phi_2\Phi_1(A). \tag{4.45}$$

(iv) Each of the maps Φ_1, Φ_2, Φ_3 is completely positive; Φ_1, Φ_2, and $\Phi_3\Phi_2\Phi_1$ are trace preserving.

The next theorem says that taking partial traces of A, B reduces the relative entropy $S(A|B)$.

4.3.12 Theorem

Let A, B be density matrices on $\mathcal{H}_1 \otimes \mathcal{H}_2$. Then

$$S(\mathrm{tr}_{\mathcal{H}_2} A|\mathrm{tr}_{\mathcal{H}_2} B) \leq S(A|B). \tag{4.46}$$

Proof. It is clear from the definition (4.29) that

$$S(U^*AU|U^*BU) = S(A|B) \tag{4.47}$$

for every unitary matrix U. Since $S(A|B)$ is jointly convex in A, B by Theorem 4.3.5, it follows from the representations (4.40) and (4.42) that

$$S(\Phi_2\Phi_1(A)|\Phi_2\Phi_1(B)) \;\leq\; S(\Phi_1(A)|\Phi_1(B)) \;\leq\; S(A|B).$$

Now note that $\Phi_2\Phi_1(A)$ is a matrix of the form $\frac{1}{m}[[\alpha_{ij}]] \otimes I_m$ and Φ_3 maps it to the $n \times n$ matrix $[[\alpha_{ij}]]$. Thus

$$S\left(\Phi_3\Phi_2\Phi_1(A)|\Phi_3\Phi_2\Phi_1(B)\right) \;=\; S\left(\Phi_2\Phi_1(A)|\Phi_2\Phi_1(B)\right).$$

This proves the theorem. ■

4.3.13 Exercise

Let $\Phi : \mathbb{M}_n \to \mathbb{M}_k$ be any completely positive trace-preserving map. Use Stinespring's theorem (Theorem 3.1.2) to show that

$$S\left(\Phi(A)|\Phi(B)\right) \;\leq\; S(A|B). \tag{4.48}$$

The inequality (4.46) is a special case of this.

Now we can state and prove the major result of this section: the *strong subadditivity* of the entropy function $S(A)$. This is a much deeper property than the subadditivity property (4.36). It is convenient to adopt some notations. We have three Hilbert spaces $\mathcal{H}_1$, $\mathcal{H}_2$, $\mathcal{H}_3$; A_{123} stands for a density matrix on $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$. A partial trace like $\operatorname{tr}_{\mathcal{H}_3} A_{123}$ is denoted as A_{12}, and so on for other indices. Likewise a partial trace $\operatorname{tr}_{\mathcal{H}_1} A_{12}$ is denoted by A_2.

4.3.14 Theorem (Lieb-Ruskai)

Let A_{123} be any density matrix in $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$. Then

$$S(A_{123}) + S(A_2) \;\leq\; S(A_{12}) \;+\; S(A_{23}). \tag{4.49}$$

Proof. By Theorem 4.3.12, taking the partial trace $\operatorname{tr}_{\mathcal{H}_3}$ gives

$$S(A_{12}|A_1 \otimes A_2) \leq S(A_{123}|A_1 \otimes A_{23}).$$

The equation (4.37) gives

$$S(A_{123}|A_1 \otimes A_{23}) \;=\; -S(A_{123}) + S(A_1) + S(A_{23}),$$

and

$$S(A_{12}|A_1 \otimes A_2) \;=\; -S(A_{12}) + S(A_1) + S(A_2).$$

Together, these three relations give (4.49). ■

4.4 FURUTA'S INEQUALITY

We have seen that the function $f(A) = A^2$ is not order preserving on positive matrices; i.e., we may have A and B for which $A \geq B \geq O$ but not $A^2 \geq B^2$. Can some weaker form of monotonicity be retrieved for the square function? This question can have different meanings. For example, we can ask whether the condition $A \geq B \geq O$ leads to an operator inequality implied by $A^2 \geq B^2$. For positive matrices A and B consider the statements

(i) $A^2 \geq B^2$;

(ii) $BA^2B \geq B^4$;

(iii) $(BA^2B)^{1/2} \geq B^2$.

Clearly (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii). Let $A \geq B$. We know that the inequality (i) does not always hold in this case. Nor does the weaker inequality (ii). A counterexample is provided by

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix}, \;\; B \;=\; \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

The matrix $BA^2B - B^4 = \begin{bmatrix} 24 & 13 \\ 13 & 7 \end{bmatrix}$ has determinant -1.

It turns out that the inequality (iii) does follow from $A \geq B$.

Note also that $A^2 \geq B^2 \Rightarrow A^4 \geq AB^2A \Rightarrow A^2 \geq (AB^2A)^{1/2}$. Once again, for A, B in the example given above we do not have $A^4 \geq AB^2A$. But we will see that $A \geq B$ always implies $A^2 \geq (AB^2A)^{1/2}$.

The most general result inspired by these considerations was proved by T. Furuta. To put it in perspective, consider the following statements:

(i) $A \geq B \geq O$;

(ii) $A^p \geq B^p,\ 0 \leq p \leq 1$;

(iii) $B^r A^p B^r \geq B^{p+2r},\ 0 \leq p \leq 1,\ r \geq 0$;

(iv) $(B^r A^p B^r)^{1/q} \geq (B^{p+2r})^{1/q},\ 0 \leq p \leq 1,\ r \geq 0,\ q \geq 1$.

We know that (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv). When $p > 1$, the implication (i) $\Rightarrow$ (ii) breaks down. Furuta's inequality says that (iv) is still valid for all p but with some restriction on q.

4.4.1 Theorem (Furuta's Inequality)

Let $A \geq B \geq O$. Then

$$(B^r A^p B^r)^{1/q} \geq B^{(p+2r)/q} \tag{4.50}$$

for $p \geq 0, r \geq 0, q \geq 1, q \geq \frac{p+2r}{1+2r}$.

Proof. For $0 \leq p \leq 1$, the inequality (4.50) is true even without the last restriction on q. So assume $p \geq 1$. If $q \geq \frac{p+2r}{1+2r}$, then $\frac{p+2r}{(1+2r)q} \leq 1$. So, the inequality (4.50) holds for such q provided it holds in the special case $q = \frac{p+2r}{1+2r}$. Thus we need to prove

$$(B^r A^p B^r)^{(1+2r)/(p+2r)} \geq B^{1+2r} \tag{4.51}$$

for $p \geq 1, r \geq 0$. We may assume that A, B are strictly positive. Let

$$B^r A^{p/2} = UP. \tag{4.52}$$

be the polar decomposition. Then

$$B^r A^p B^r = UP^2\, U^*.$$

Hence, for any $q > 0$

$$(B^r A^p B^r)^{1/q} = UP^{2/q} U^*,$$

and, therefore (using (4.52) thrice) we get

$$\begin{aligned} B^{-r}(B^rA^pB^r)^{1/q}B^{-r} &= B^{-r}UP^{2/q}U^*B^{-r} \\ &= (A^{p/2}P^{-1}U^*)(UP^{2/q}U^*)(UP^{-1}A^{p/2}) \\ &= A^{p/2}(P^2)^{1/q-1}A^{p/2} \\ &= A^{p/2}(A^{p/2}B^{2r}A^{p/2})^{1/q-1}A^{p/2} \\ &= A^{p/2}(A^{-p/2}B^{-2r}A^{-p/2})^{1-1/q}A^{p/2}. \end{aligned} \tag{4.53}$$

Now suppose $0 \le r \le 1/2$. Then $A^{2r} \ge B^{2r}$, and hence $B^{-2r} \ge A^{-2r}$. Choose $q = \frac{p+2r}{1+2r}$. Then $1 - \frac{1}{q} = \frac{p-1}{p+2r} \le 1$. So, we get from (4.53)

$$B^{-r}(B^rA^pB^r)^{1/q}B^{-r} \ge A^{p/2}(A^{-p/2}A^{-2r}A^{-p/2})^{(p-1)/(p+2r)}A^{p/2} = A.$$

Thus

$$(B^rA^pB^r)^{1/q} \ge B^rAB^r \ge B^{1+2r}. \tag{4.54}$$

We have thus proved the inequality (4.51) for r in the domain $[0, 1/2]$. This domain is extended by inductive steps as follows. Let

$$A_1 = (B^rA^pB^r)^{1/q},\ B_1 = B^{1+2r}, \tag{4.55}$$

where $r \in [0, 1/2]$ and $q = (p+2r)/(1+2r)$. We have proved that $A_1 \ge B_1$. Let p_1, r_1 be any numbers with $p_1 \ge 1$, $r_1 \in [0, 1/2]$ and let $q_1 = (p_1 + 2r_1)/(1+2r_1)$. Apply the inequality (4.54) to A_1, B_1, p_1, r_1, q_1 to get

$$(B_1^{r_1}A_1^{p_1}B_1^{r_1})^{1/q_1} \ \ge\ B_1^{1+2r_1}. \tag{4.56}$$

This is true, in particular, when $p_1 = q$ and $r_1 = 1/2$. So we have

$$(B_1^{1/2}A_1^qB_1^{1/2})^{1/q_1} \ge B_1^2.$$

Substitute the values of A_1, B_1 from (4.55) to get from this

$$(B^{2r+1/2}A^pB^{2r+1/2})^{1/q_1} \ \ge\ B^{2(1+2r)}. \tag{4.57}$$

Put $2r + 1/2 = s$, and note that with the choices just made

$$\begin{aligned} q_1 = \frac{p_1 + 2r_1}{1 + 2r_1} = \frac{q+1}{2} = \frac{(p+2r)/(1+2r)+1}{2} \\ = \frac{p + 4r + 1}{4r + 2} = \frac{p+2s}{1+2s}. \end{aligned}$$

So, (4.57) can be written as

$$(B^s A^p B^s)^{(1+2s)/(p+2s)} \geq B^{1+2s},$$

where $s \in [1/2, 3/2]$. Thus we have enlarged the domain of validity of the inequality (4.51) from r in $[0, 1/2]$ to r in $[0, 3/2]$. The process can be repeated to see that the inequality is valid for all $r \geq 0$. ■

4.4.2 Corollary

Let A, B, p, q, r be as in the Theorem. Then

$$(A^{p+2r})^{1/q} \geq (A^r B^p A^r)^{1/q}. \tag{4.58}$$

Proof. Assume A, B are strictly positive. Since $B^{-1} \geq A^{-1} > O$, (4.50) gives us

$$(A^{-r} B^{-p} A^{-r})^{1/q} \geq A^{-(p+2r)/q}.$$

Taking inverse on both sides reverses this inequality and gives us (4.58). ■

4.4.3 Corollary

Let $A \geq B \geq O$, $p \geq 1$, $r \geq 0$. Then

$$(B^r A^p B^r)^{1/p} \geq (B^{p+2r})^{1/p}, \tag{4.59}$$

$$(A^{p+2r})^{1/p} \geq (A^r B^p A^r)^{1/p}. \tag{4.60}$$

4.4.4 Corollary

Let $A \geq B \geq O$. Then

$$(BA^2B)^{1/2} \geq B^2, \tag{4.61}$$

$$A^2 \geq (AB^2A)^{1/2}. \tag{4.62}$$

These are the statements with which we began our discussion in this section. Another special consequence of Furuta's inequality is the following.

4.4.5 Corollary

Let $A \geq B \geq O$. Then for $0 < p < \infty$

$$A^p \# B^{-p} \geq I. \tag{4.63}$$

Proof. Choose $r = p/2$ and $q = 2$ in (4.58) to get

$$A^p \geq \left(A^{p/2} B^p A^{p/2}\right)^{1/2}.$$

This is equivalent to the inequality

$$I \geq A^{-p/2} \left(A^{p/2} B^p A^{p/2}\right)^{1/2} A^{-p/2} = A^{-p} \# B^p.$$

Using the relation (4.17) we get (4.63). ■

For $0 \leq p \leq 1$, the inequality (4.63) follows from Theorem 4.2.1. While the theorem does need this restriction on p, the inequality (4.63) exhibits a weaker monotonicity of the powers A^p for $p > 1$.

4.5 SUPPLEMENTARY RESULTS AND EXERCISES

The matrix equation

$$AX - XB = Y \tag{4.64}$$

is called the Sylvester equation. If no eigenvalue of A is an eigenvalue of B, then for every Y this equation has a unique solution X. The following exercise outlines a proof of this.

4.5.1 Exercise

(i) Let $\mathcal{A}(X) = AX$. This is a linear operator on $\mathbb{M}_n$. Each eigenvalue of A is an eigenvalue of $\mathcal{A}$ with multiplicity n times as much. Likewise the eigenvalues of the operator $\mathcal{B}(X) = XB$ are the eigenvalues of B.

(ii) The operators $\mathcal{A}$ and $\mathcal{B}$ commute. So the spectrum of $\mathcal{A} - \mathcal{B}$ is contained in the difference $\sigma(\mathcal{A}) - \sigma(\mathcal{B})$, where $\sigma(\mathcal{A})$ stands for the spectrum of $\mathcal{A}$.

(iii) Thus if $\sigma(A)$ and $\sigma(B)$ are disjoint, then $\sigma(\mathcal{A} - \mathcal{B})$ does not contain the point 0. Hence the operator $\mathcal{A} - \mathcal{B}$ is invertible. This is the same as saying that for each Y, there exists a unique X satisfying the equation (4.64).

The Lyapunov equation (1.14) is a special type of Sylvester equation.

There are various ways in which functions of an arbitrary matrix may be defined. One standard approach via the Jordan canonical form tells us how to explicitly write down a formula for $f(A)$ for every function that is $n-1$ times differentiable on an open set containing $\sigma(A)$. Using this one can see that if A is a matrix whose spectrum is in $(0, \infty)$, then it has a square root whose spectrum is also in $(0, \infty)$. Another standard approach using power series expansions is equally useful.

4.5.2 Exercise

Let $B_1^2 = A$ and $B_2^2 = A$. Then

$$B_1(B_1 - B_2) + (B_1 - B_2)B_2 = O.$$

Suppose all eigenvalues of B_1 and B_2 are positive. Use the (uniqueness part of) Exercise 4.5.1 to show that $B_1 = B_2$. This shows that for every matrix A whose spectrum is contained in $(0, \infty)$ there is a unique matrix B for which $B^2 = A$ and $\sigma(B)$ is contained in $(0, \infty)$.

The same argument shows that if $\sigma(A)$ is contained in the open right half plane, then there is a unique matrix B with the same property that satisfies the equation $B^2 = A$.

4.5.3 Exercise

Let Φ be a positive unital linear map on $\mathbb{M}_n$. Use Theorem 4.1.5(i) to show that if A and B are strictly positive, then

$$\left(\Phi\left((A+B)^{-1}\right)\right)^{-1} \geq \left(\Phi(A^{-1})\right)^{-1} + \left(\Phi(B^{-1})\right)^{-1}.$$

Thus the map $A \mapsto \left(\Phi(A^{-1})\right)^{-1}$ is monotone and concave on the set of positive matrices.

4.5.4 Exercise

Let Φ be a positive unital linear map. Show that for all positive matrices A

$$\log\, \Phi(A) \geq \Phi(\log A).$$

(See Proposition 2.7.1.)

The Schur product $A \circ B$ is a principal submatrix of $A \otimes B$. So, there is a positive unital linear map Φ from $\mathbb{M}_{n^2}$ into $\mathbb{M}_n$ such that $\Phi(A \otimes B) = A \circ B$. This observation is useful in deriving convexity and concavity results about Schur products from those about tensor products. We leave it to the reader to obtain such results from what we have done in Chapters 2 and 4.

The arithmetic, geometric, and harmonic means are the best-known examples of means. We have briefly alluded to the logarithmic mean in (4.2). Several other means arise in various contexts. We mention two families of such means.

For $0 \leq \nu \leq 1$ let

$$H_\nu(a,b) = \frac{a^\nu\, b^{1-\nu} + a^{1-\nu}\, b^\nu}{2}. \tag{4.65}$$

We call these the *Heinz means.* Notice that $H_\nu = H_{1-\nu}$, $H_{1/2}$ is the geometric mean, and $H_0 = H_1$ is the arithmetic mean. Thus, the family H_ν interpolates between the geometric and the arithmetic mean. Each H_ν satisfies conditions (i)–(vi) for means.

4.5.5 Exercise

(i) For fixed positive numbers a and b, $H_\nu(a,b)$ is a convex function of ν in the interval $[0,1]$, and attains its minimum at $\nu = 1/2$.

Thus

$$\sqrt{ab} \le H_\nu(a,b) \le \frac{a+b}{2}. \tag{4.66}$$

(ii) Show that

$$\int_0^1 H_\nu(a,b)d\nu = L(a,b), \tag{4.67}$$

and use this to prove the inequality (4.3).

For $-\infty \le p \le \infty$ let

$$B_p(a,b) = \Big(\frac{a^p+b^p}{2}\Big)^{1/p}. \tag{4.68}$$

These are called the *power means* or the *binomial means.* Here it is understood that

$$\begin{aligned} B_0(a,b) &= \lim_{p\to 0} B_p(a,b) = \sqrt{ab}, \\ B_\infty(a,b) &= \lim_{p\to\infty} B_p(a,b) = \max(a,b), \\ B_{-\infty}(a,b) &= \lim_{p\to-\infty} B_p(a,b) = \min(a,b). \end{aligned}$$

The arithmetic and the harmonic means are included in this family. Properties (i)–(vi) of means may readily be verified for this family.

In Section 4.1 we defined the geometric mean $A\#B$ by using the congruence $\Gamma_{A^{-1/2}}$ to reduce the pair (A,B) to the commuting pair $(I, A^{-1/2}BA^{-1/2})$. A similar procedure may be used for the other means. Given a mean M on positive numbers, let

$$f(x) = M(x,1). \tag{4.69}$$

4.5.6 Exercise

From properties (i)–(vi) of M deduce that the function f on $\mathbb{R}_+$ has the following properties:

(i) $f(1) = 1$,

(ii) $xf(x^{-1}) = f(x)$,

(iii) f is monotone increasing,

(iv) f is continuous,

(v) $f(x) \le 1$ for $0 < x \le 1$, and $f(x) \ge 1$ for $x \ge 1$.

4.5.7 Exercise

Let f be a map of $\mathbb{R}_+$ into itself satisfying the properties (i)–(v) given in Exercise 4.5.6. For positive numbers a and b let

$$M(a, b) = a\, f(b/a). \tag{4.70}$$

Show that M is a mean.

Given a mean $M(a, b)$ on positive numbers let $f(x)$ be the function associated with it by the relation (4.69). For positive matrices A and B let

$$M(A, B) = A^{1/2}\; f(A^{-1/2}BA^{-1/2})A^{1/2}. \tag{4.71}$$

When $M(a, b) = \sqrt{ab}$ this formula gives the geometric mean $A\#B$ defined in (4.10). Does this procedure always lead to an operator mean satisfying conditions (i)–(vi)? For the geometric mean we verified its symmetry by an indirect argument. The next proposition says that such symmetry is a general fact.

4.5.8 Proposition

Let $M(A, B)$ be defined by (4.71). *Then for all A* and *B*

$$M(A, B) = M(B, A). \tag{4.72}$$

Proof. We have to show that

$$f(A^{-1/2}BA^{-1/2}) = A^{-1/2}B^{1/2}f(B^{-1/2}AB^{-1/2})B^{1/2}A^{-1/2}. \tag{4.73}$$

If $A^{-1/2}B^{1/2} = UP$ is the polar decomposition, then $B^{1/2}A^{-1/2} = PU^*$ and $B^{-1/2}A^{1/2} = P^{-1}U^*$. The left-hand side of (4.73) is, therefore, equal to

$$f(UP^2U^*) = Uf(P^2)U^*.$$

The right-hand side of (4.73) is equal to

$$UPf(P^{-2})PU^*.$$

So, (4.73) will be proved if we show that $f(P^2) = Pf(P^{-2})P$. This follows from the fact that for every x we have $f(x^2) = x^2 f(x^{-2})$ as shown in Exercise 4.5.6. ■

4.5.9 Exercise

Show that $M(A, B)$ defined by (4.71) is invariant under congruence; i.e.,

$$M(X^*AX,\ X^*BX) = X^*\ M(A, B)X$$

for every invertible matrix X. [Hint: Use the polar decomposition.]

Our next concern is whether $M(A, B)$ defined by (4.71) is monotone in the variables A and B. This is so *provided the function f is operator monotone.* In this case monotonicity in B is evident from the formula (4.71). By symmetry it is monotone in A as well.

For the means that we have considered in this chapter the function f is given by

$$\begin{aligned}
f(x) &= \frac{1+x}{2} \quad \text{(arithmetic mean)},\\
f(x) &= \sqrt{x} \quad \text{(geometric mean)},\\
f(x) &= \frac{2x}{1+x} \quad \text{(harmonic mean)},\\
f(x) &= \int_0^1 x^t dt \quad \text{(logarithmic mean)},\\
f(x) &= \frac{x^\nu + x^{1-\nu}}{2},\ \ 0 \le \nu \le 1 \quad \text{(Heinz means)},\\
f(x) &= \left(\frac{x^p+1}{2}\right)^{1/p},\ \ -\infty \le p \le \infty \quad \text{(binomial means)}.
\end{aligned}$$

The first five functions in this list are operator monotone. The last enjoys this property only for some values of p.

4.5.10 Exercise

Let f be an operator monotone function on $(0, \infty)$. Then the function $g(x) = [f(x^p)]^{1/p}$ is operator monotone for $0 < p \le 1$. [It may be helpful, in proving this, to use a theorem of Loewner that says f is operator monotone if and only if it has an analytic continuation mapping the upper half-plane into itself. See MA Chapter V.]

4.5.11 Exercise

Show that the function

$$f(x) = \left(\frac{x^p+1}{2}\right)^{1/p}$$

is operator monotone if and only if $-1 \le p \le 1$.

Thus the binomial means $B_p(a,b)$ defined by (4.68) lead to matrix means satisfying all our requirements if and only if $-1 \le p \le 1$. The mean $B_0(a,b)$ leads to the geometric mean $A\#B$.

The logarithmic mean is important in different contexts, one of them being heat flow. We explain this briefly. Heat transfer by steady unidirectional conduction is governed by *Fourier's law.* If the direction of heat flow is along the x-axis, this law says

$$q = kA\frac{dT}{dx}, \tag{4.74}$$

where q is the rate of heat flow along the x-axis across an area A normal to the x-axis, dT/dx is the temperature gradient along the x direction, and k is a constant called thermal conductivity of the material through which the heat is flowing.

The cross-sectional area A may be constant, as for example in a cube. More often, as in the case of a fluid traveling in a pipe, it is a variable. In such cases it is convenient for engineering calculations to write (4.74) as

$$q = kA_m\frac{\Delta T}{\Delta x} \tag{4.75}$$

where A_m is the *mean cross section* of the body between two points at distance Δx along the x-axis and ΔT is the difference of temperatures at these two points. For example, in the case of a body with uniformly tapering rectangular cross section, A_m is the arithmetic mean of the two boundary areas A_1 and A_2.

A very common situation is that of a liquid flowing through a long hollow cylinder (like a pipe). Here heat flows through the sides of the cylinder in a radial direction perpendicular to the axis of the cylinder. The cross sectional area in this case is proportional to the distance from the centre.

Consider the section of the cylinder bounded by two concentric cylinders at distances x_1 and x_2 from the centre. Total heat flow across this section given by (4.74) is

$$\int_{x_1}^{x_2}\frac{dx}{A} = k\frac{\Delta T}{q}, \tag{4.76}$$

where $A = 2\pi x L$, L being the length of the cylinder. This shows that

$$q = \frac{k\ 2\pi L\ \Delta T}{\log\, x_2 - \log x_1}.$$

If we wish to write it in the form (4.75) with $\Delta x = x_2 - x_1$, then we must have

$$A_m = 2\pi L \frac{x_2 - x_1}{\log x_2 - \log x_1} = \frac{2\pi L x_2 - 2\pi L x_1}{\log 2\pi L x_2 - \log 2\pi L x_1}.$$

In other words,

$$A_m = \frac{A_2 - A_1}{\log A_2 - \log A_1},$$

the logarithmic mean of the two areas bounding the section under consideration. In the chemical engineering literature this is called the *logarithmic mean area.*

If instead of a hollow cylinder we consider a hollow sphere, then the cross-sectional area is proportional to the square of the distance from the center. In this case we get from (4.76)

$$\int_{x_1}^{x_2} \frac{dx}{4\pi x^2} = k \frac{\Delta T}{q}.$$

The reader can check that in this case

$$A_m = \sqrt{A_1 A_2},$$

the geometric mean of the two areas bounding the annular section under consideration.

In Chapter 6 we will see that the inequality between the geometric and the logarithmic mean plays a fundamental role in differential geometry.

4.6 NOTES AND REFERENCES

The parallel sum (4.4) was introduced by W. N. Anderson and R. J. Duffin, *Series and parallel addition of matrices,* J. Math. Anal. Appl., 26 (1969) 576–594. They also proved some of the basic properties of this object like monotonicity and concavity, and the arithmetic-harmonic mean inequality. Several other operations corresponding to different kinds of electrical networks were defined following the parallel sum. See W. N. Anderson and G. E. Trapp, *Matrix operations induced by electrical network connections—a survey,* in *Constructive Approaches to Mathematical Models,* Academic Press, 1979, pp. 53–73.

The geometric mean formula (4.10) and some of its properties like (4.13) are given in W. Pusz and S. L. Woronowicz, *Functional calculus* *for sesquilinear forms and the purification map,* Rep. Math. Phys., 8 (1975) 159–170. The notations and the language of this paper are different from ours. It was in the beautiful paper T. Ando, *Concavity of certain maps on positive definite matrices and applications to Hadamard products,* Linear Algebra Appl., 26 (1979) 203–241, that several concepts were clearly stated, many basic properties demonstrated, and the power of the idea illustrated through several applications. The interplay between means and positive linear maps clearly comes out in this paper, concavity and convexity of various maps are studied, a new proof of Lieb's theorem 4.27 is given and many new kinds of inequalities for the Schur product are obtained. This paper gave a lot of impetus to the study of matrix means, and of matrix inequalities in general.

W. N. Anderson and G. E. Trapp, *Operator means and electrical networks,* in Proc. 1980 IEEE International Symposium on Circuits and Systems, pointed out that $A\#B$ is the positive solution of the Riccati equation (4.11). They also drew attention to the matrix in Proposition 4.1.9. This had been studied in H. J. Carlin and G. A. Noble, *Circuit properties of coupled dispersive lines with applications to wave guide modelling,* in *Proc. Network and Signal Theory,* J. K. Skwirzynski and J. O. Scanlan, eds., Peter Pergrinus, 1973, pp. 258–269. Note that this paper predates the one by Pusz and Woronowicz.

An axiomatic theory of matrix means was developed in F. Kubo and T. Ando, *Means of positive linear operators,* Math. Ann., 246 (1980) 205–224. Here it is shown that there is a one-to-one correspondence between matrix monotone functions and matrix means. This is implicit in some of our discussion in Section 4.5.

A notion of geometric mean different from ours is introduced and studied by M. Fiedler and V. Pták, *A new positive definite geometric mean of two positive definite matrices,* Linear Algebra Appl., 251 (1997) 1–20. This paper contains a discussion of the mean $A\#B$ as well.

Entropy is an important notion in statistical mechanics and in information theory. Both of these subjects have their classical and quantum versions. Eminently readable introductions are given by E. H. Lieb and J. Yngvason, *A guide to entropy and the second law of thermodynamics,* Notices Am. Math. Soc., 45 (1998) 571–581, and in other articles by these two authors such as *The mathematical structure of the second law of thermodynamics,* in *Current Developments in Mathematics, 2001,* International Press, 2002. The two articles by A. Wehrl, *General properties of entropy,* Rev. Mod. Phys., 50 (1978) 221–260, and *The many facets of entropy,* Rep. Math. Phys., 30 (1991) 119–129, are comprehensive surveys of various aspects of the subject. The text by M. Ohya and D. Petz, *Quantum Entropy and Its Use,* Springer, 1993, is another resource. The book by M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information,* Cambridge University Press, 2000, reflects the renewed vigorous interest in this topic because of the possibility of new significant applications to information technology.

Let $(p_1, \ldots, p_n)$ be a probability vector; i.e., $p_i \geq 0$ and $\sum p_i = 1$. C. Shannon, *Mathematical theory of communication,* Bell Syst. Tech. J., 27 (1948) 379–423, introduced the function $S(p_1, \ldots, p_n) = -\sum p_i \log p_i$ as a measure of "average lack of information" in a statistical experiment with n possible outcomes occurring with probabilities $p_1, \ldots, p_n$. The quantum analogue of a probability vector is a density matrix A; i.e., $A \geq O$ and $\operatorname{tr} A = 1$. The quantum entropy function $S(A) = -\operatorname{tr} A \log A$ was defined by J. von Neumann, *Thermodynamik quantenmechanischer Gesamtheiten,* Göttingen Nachr., 1927, pp. 273–291; see also Chapter 5 of his book *Mathematical Foundations of Quantum Mechanics,* Princeton University Press, 1955. Thus von Neumann's definition preceded Shannon's. There were strong motivations for the former because of the work of nineteenth-century physicists, especially L. Boltzmann.

Theorem 4.3.3 was proved in E. H. Lieb, *Convex trace functions and the Wigner-Yanase-Dyson conjecture,* Adv. Math., 11 (1973) 267–288. Because of its fundamental interest and importance, several different proofs appeared soon after Lieb's paper. One immediate major application of this theorem was made in the proof of Theorem 4.3.14 by E. H. Lieb and M. B. Ruskai, *A fundamental property of quantum-mechanical entropy,* Phys. Rev. Lett., 30 (1973) 434–436, and *Proof of the strong subadditivity of quantum-mechanical entropy,* J. Math. Phys., 14 (1973) 1938–1941. Several papers of Lieb are conveniently collected in *Inequalities, Selecta of Elliot H. Lieb,* M. Loss and M. B. Ruskai eds., Springer, 2002. The matrix-friendly proof of Theorem 4.3.14 is adopted from R. Bhatia, *Partial traces and entropy inequalities,* Linear Algebra Appl., 375 (2003) 211–220.

Three papers of G. Lindblad, *Entropy, information and quantum measurements,* Commun. Math. Phys., 33 (1973) 305–322, *Expectations and entropy inequalities for finite quantum systems,* ibid. 39 (1974) 111–119, and *Completely positive maps and entropy inequalities,* ibid. 40 (1975) 147–151, explore various convexity properties of entropy and their interrelations. For example, the equivalence of Theorems 4.3.5 and 4.3.14 is noted in the second paper and the inequality (4.48) is proved in the third.

In the context of quantum statistical mechanics the tensor product operation represents the physical notion of putting a system in a larger system. In quantum information theory it is used to represent the notion of entanglement of states. These considerations have led to several problems very similar to the ones discussed in Section 4.3. We mention one of these as an illustrative example. Let Φ be a completely positive trace-preserving (CPTP) map on $\mathbb{M}_n$. The *minimal entropy* of the "quantum channel" Φ is defined as

$$S_{\min}(\Phi) = \inf \{S(\Phi(A)) : A \geq O,\ \ \operatorname{tr} A = 1\}.$$

It is conjectured that

$$S_{\min}(\Phi_1 \otimes \Phi_2) = S_{\min}(\Phi_1) + S_{\min}(\Phi_2)$$

for any two CPTP maps Φ_1 and Φ_2. See P. W. Shor, *Equivalence of additivity questions in quantum information theory,* Commun. Math. Phys., 246 (2004) 453–472 for a statement of several problems of this type and their importance.

Furuta's inequality was proved by T. Furuta, $A \geq B \geq O$ *assures* $(B^rA^pB^r)^{1/q} \geq B^{(p+2r)/q}$ *for* $r \geq 0, p \geq 0, q \geq 1$ *with* $(1+2r)q \geq p+2r$, Proc. Am. Math. Soc. 101 (1987) 85–88. This paper sparked off several others giving different proofs, extensions, and applications. For example, T. Ando, *On some operator inequalities,* Math. Ann., 279 (1987) 157–159, showed that if $A \geq B$, then $e^{-tA} \# e^{tB} \leq I$ for all $t \geq 0$. It was pointed out at the beginning of Section 4.4 that $A \geq B \geq O$ does not imply $A^2 \geq B^2$ but it does imply the weaker inequality $(BA^2B)^{1/2} \geq B^2$. This can be restated as $A^{-2} \# B^2 \leq I$. In a similar vein, $A \geq B$ does not imply $e^A \geq e^B$ but it does imply $e^{-A} \# e^B \leq I$. It is not surprising that Furuta's inequality is related to the theory of means and to properties of matrix exponential and logarithm functions.

The name "Heinz means" for (4.65) is not standard usage. We have called them so because of the famous inequalities of E. Heinz (proved in Chapter 5). The means (4.68) have been studied extensively. See for example, G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities,* Second Edition, Cambridge University Press, 1952. The matrix analogues

$$M_p(A, B) = \left(\frac{A^p + B^p}{2}\right)^{1/p}$$

are analysed in K. V. Bhagwat and R. Subramanian, *Inequalities between means of positive operators,* Math. Proc. Cambridge Philos.

Soc., 83 (1978) 393–401. It is noted there that the limiting value $M_0(A,B) = \exp\left(\frac{\log A + \log B}{2}\right)$. We have seen that this is not a suitable definition of the geometric mean of A and B.

A very interesting article on the matrix geometric mean is J. D. Lawson and Y. Lim, *The geometric mean, matrices, metrics, and more,* Am. Math. Monthly 108 (2001) 797–812. The importance of the logarithmic mean in engineering problems is discussed, for example, in W. H. McAdams, *Heat Transmission,* Third Edition, McGraw Hill, 1954.

Chapter Five

Positive Definite Functions

Positive definite functions arise naturally in many areas of mathematics. In this chapter we study some of their basic properties, construct some examples, and use them to derive interesting results about positive matrices.

5.1 BASIC PROPERTIES

Positive definite sequences were introduced in Section 1.1.3. We repeat the definition. A (doubly infinite) sequence of complex numbers $\{a_n : n \in \mathbb{Z}\}$ is said to be *positive definite* if for every positive integer N, we have

$$\sum_{r,s=0}^{N-1} a_{r-s}\, \xi_r\, \overline{\xi}_s \geq 0, \tag{5.1}$$

for every finite sequence of complex numbers $\xi_0,\ \xi_1, \ldots, \xi_{N-1}$. This condition is equivalent to the requirement that for each $N = 1, 2, \ldots,$ the $N \times N$ matrix

$$\begin{bmatrix} a_0 & a_{-1} & \cdots & a_{-(N-1)} \\ a_1 & a_0 & a_{-1} & \cdots \\ \vdots & & \cdots & a_{-1} \\ a_{N-1} & \cdots & a_1 & a_0 \end{bmatrix} \tag{5.2}$$

is positive.

From this condition it is clear that we must have

$$a_0 \geq 0, \quad a_{-n} = \overline{a}_n, \quad |a_n| \leq a_0. \tag{5.3}$$

A complex-valued function φ on $\mathbb{R}$ is said to be *positive definite* if for every positive integer N, we have

$$\sum_{r,s=0}^{N-1} \varphi(x_r - x_s)\, \xi_r\, \overline{\xi}_s \geq 0, \tag{5.4}$$

for every choice of real numbers $x_0,\ x_1,\ldots,\ x_{N-1}$, and complex numbers $\xi_0,\ \xi_1,\ldots,\ \xi_{N-1}$. In other words φ is positive definite if for each $N = 1, 2, \ldots$ the $N \times N$ matrix

$$[[\varphi(x_r - x_s)]] \tag{5.5}$$

is positive for every choice of real numbers $x_0,\ \ldots,\ x_{N-1}$. It follows from this condition that

$$\varphi(0) \geq 0, \quad \varphi(-x) = \overline{\varphi(x)}, \quad |\varphi(x)| \leq \varphi(0). \tag{5.6}$$

Thus every positive definite function is bounded, and its maximum absolute value is attained at 0.

5.1.1 Exercise

Let $\varphi(x)$ be the characteristic function of the set $\mathbb{Z}$; i.e., $\varphi(x) = 1$ if $x \in \mathbb{Z}$ and $\varphi(x) = 0$ if $x \in \mathbb{R} \setminus \mathbb{Z}$. Show that φ is positive definite. This remains true when $\mathbb{Z}$ is replaced by any additive subgroup of $\mathbb{R}$.

5.1.2 Lemma

If φ is positive definite, then for all $x_1,\ x_2$

$$|\varphi(x_1) - \varphi(x_2)|^2 \leq 2\,\varphi(0)\,\mathrm{Re}[\varphi(0) - \varphi(x_1 - x_2)].$$

Proof. Assume, without loss of generality, that $\varphi(0) = 1$. Choose $x_0 = 0$. The 3×3 matrix

$$A = \begin{bmatrix} 1 & \overline{\varphi(x_1)} & \overline{\varphi(x_2)} \\ \varphi(x_1) & 1 & \varphi(x_1 - x_2) \\ \varphi(x_2) & \overline{\varphi(x_1 - x_2)} & 1 \end{bmatrix}$$

is positive. So for every vector u, the inner product $\langle u, Au\rangle \geq 0$. Choose $u = (z, 1, -1)$ where z is any complex number. This gives the inequality

$$-2\,\mathrm{Re}\{z(\varphi(x_1) - \varphi(x_2))\} - |z|^2 \leq 2\,[1 - \mathrm{Re}\,\varphi(x_1 - x_2)].$$

Now choose $z = \overline{\varphi(x_2) - \varphi(x_1)}$. This gives

$$|\varphi(x_2) - \varphi(x_1)|^2 \leq 2\,\mathrm{Re}\,[1 - \varphi(x_1 - x_2)]. \qquad \blacksquare$$

Exercise 5.1.1 showed us that a positive definite function φ need not be continuous. Lemma 5.1.2 shows that if the real part of φ is continuous at 0, then φ is continuous everywhere on $\mathbb{R}$.

5.1.3 Exercise

Let $\varphi(x)$ be positive definite. Then

(i) $\overline{\varphi(x)}$ is positive definite.

(ii) For every real number t the function $\varphi(tx)$ is positive definite.

5.1.4 Exercise

(i) If φ_1, φ_2 are positive definite, then so is their product $\varphi_1\varphi_2$. (Schur's theorem.)

(ii) If φ is positive definite, then $|\varphi|^2$ is positive definite. So is $\mathrm{Re}\,\varphi$.

5.1.5 Exercise

(i) If $\varphi_1, \ldots, \varphi_n$ are positive definite, and $a_1, \ldots, a_n$ are positive real numbers, then $a_1\varphi_1 + \cdots + a_n\varphi_n$ is positive definite.

(ii) If $\{\varphi_n\}$ is a sequence of positive definite functions and $\varphi_n(x) \to \varphi(x)$ for all x, then φ is positive definite.

(iii) If φ is positive definite, then e^{φ} is positive definite, and so is $e^{\varphi+a}$ for every $a \in \mathbb{R}$.

(iv) If $\varphi(x)$ is a measurable positive definite function and $f(t)$ is a nonnegative integrable function, then $\int_{-\infty}^{\infty} \varphi(tx)f(t)\, dt$ is positive definite.

(v) If μ is a finite positive Borel measure on $\mathbb{R}$ and $\varphi(x)$ a measurable positive definite function, then the function $\int_{-\infty}^{\infty} \varphi(tx)\, d\mu(t)$ is positive definite. (The statement (iv) is a special case of (v).)

Let I be any interval and let $K(x, y)$ be a bounded continuous complex-valued function on $I \times I$. We say K is a *positive definite kernel* if

$$\int_I \int_I K(x,y)\; f(x)\; \overline{f(y)}\; dx\, dy \geq 0 \tag{5.7}$$

for every continuous integrable function f on the interval I.

5.1.6 Exercise

(i) A bounded continuous function $K(x, y)$ on $I \times I$ is a positive definite kernel if and only if for all choices of points $x_1, \ldots, x_N$ in I, the $N \times N$ matrix $[[K(x_i, x_j)]]$ is positive.

(ii) A bounded continuous function φ on $\mathbb{R}$ is positive definite if and only if the kernel $K(x, y) = \varphi(x - y)$ is positive definite.

5.2 EXAMPLES

5.2.1

The function $\varphi(x) = e^{ix}$ is positive definite since

$$\sum_{r,s} e^{i(x_r - x_s)}\; \xi_r\, \overline{\xi}_s = \left| \sum_r e^{ix_r}\; \xi_r \right|^2 \geq 0.$$

Exercise: Write the matrix (5.5) in this case as uu^* where u is a column vector.

This example is fundamental. It is a remarkable fact that *all* positive definite functions can be built from this one function by procedures outlined in Section 5.1.

5.2.2

The function $\varphi(x) = \cos x$ is positive definite.

Exercise: $\sin x$ is *not* positive definite. The matrix A in Exercise 1.6.6 has entries $a_{ij} = \cos(x_i - x_j)$ where $x_i = 0,\ \pi/4,\ \pi/2,\ 3\pi/4$. It follows that $|\cos x|$ is not positive definite.

5.2.3

For each $t \in \mathbb{R}$, $\varphi(x) = e^{itx}$ is a positive definite function.

5.2.4

Let $f \in L_1(\mathbb{R})$ and let $f(t) \geq 0$. Then

$$\hat{f}(x) := \int_{-\infty}^{\infty} e^{-itx}\, f(t)\, dt \tag{5.8}$$

is positive definite. More generally, if μ is a positive finite Borel measure on $\mathbb{R}$, then

$$\hat{\mu}(x) := \int_{-\infty}^{\infty} e^{-itx}\, d\mu(t) \tag{5.9}$$

is positive definite. The function $\hat{f}$ is called the *Fourier transform* of f and $\hat{\mu}$ is called the *Fourier-Stieltjes transform* of μ. Both of them are bounded uniformly continuous functions.

These transforms give us many interesting positive definite functions.

5.2.5

One of the first calculations in probability theory is the computation of an integral:

$$\int_{-\infty}^{\infty} e^{-itx}\, e^{-t^2/2}\, dt = e^{-x^2/2} \int_{-\infty}^{\infty} e^{-(t+ix)^2/2}\, dt.$$

The integral on the right hand side can be evaluated using Cauchy's theorem. Let C be the rectangular contour with vertices $-R, R, R+ix, -R+ix$. The integral of the analytic function $f(z) = e^{-z^2/2}$ along this contour is zero. As $R \to \infty$, the integral along the two vertical sides of this contour goes to zero. Hence

$$\int_{-\infty}^{\infty} e^{-(t+ix)^2/2}\, dt = \int_{-\infty}^{\infty} e^{-t^2/2}\, dt = \sqrt{2\pi}.$$

So,

$$\int_{-\infty}^{\infty} e^{-itx} e^{-t^2/2}\, dt = \sqrt{2\pi} e^{-x^2/2}. \tag{5.10}$$

(This shows that, with a suitable normalization, the function $e^{-x^2/2}$ is its own Fourier transform.) Thus for each $a \geq 0$, the function $\varphi(x) = e^{-ax^2}$ is positive definite.

5.2.6

The function $\varphi(x) = \sin x/x$ is positive definite. To see this one can use the product formula

$$\frac{\sin x}{x} = \prod_{k=1}^{\infty} \cos \frac{x}{2^k}, \tag{5.11}$$

and observe that each of the factors in this infinite product is a positive definite function. Alternately, we can use the formula.

$$\frac{\sin x}{x} = \frac{1}{2} \int_{-1}^{1} e^{-itx}\, dt. \tag{5.12}$$

(This integral is the Fourier transform of the characteristic function $\chi_{[-1,1]}$.)

We have tacitly assumed here that $\varphi(0) = 1$. This is natural. If we assign $\varphi(0)$ any value larger than 1, the resulting (discontinuous) function is also positive definite.

5.2.7

The integral

$$\int_0^\infty e^{-itx}\, e^{-t}\, dt = \frac{1}{1+ix}$$

shows that the function $\varphi(x) = 1/(1+ix)$ is positive definite. The functions $1/(1-ix)$ and $1/(1+x^2)$ are positive definite.

5.2.8

The integral formulas

$$\frac{1}{1+x^2} = \frac{1}{2}\int_{-\infty}^\infty e^{-itx}\, e^{-|t|}\, dt$$

and

$$e^{-|x|} = \frac{1}{\pi}\int_{-\infty}^\infty \frac{e^{-itx}}{1+t^2}\, dt$$

show that the functions $1/(1+x^2)$ and $e^{-|x|}$ are positive definite. (They are nonnegative and are, up to a constant factor, Fourier transforms of each other.)

5.2.9

From the product representations

$$\frac{\sinh x}{x} = \prod_{k=1}^\infty \left(1 + \frac{x^2}{k^2\,\pi^2}\right) \tag{5.13}$$

and

$$\cosh x = \prod_{k=0}^{\infty} \left(1 + \frac{4x^2}{(2k+1)^2\,\pi^2}\right), \tag{5.14}$$

one sees (using the fact that $1/(1+a^2x^2)$ is positive definite) that the functions $x/(\sinh x)$ and $1/(\cosh x)$ are positive definite. (Contrast this with 5.2.6 and 5.2.2.)

5.2.10

For $0 < \alpha < 1$, we have from (5.13)

$$\frac{\sinh \alpha x}{\sinh x} = \alpha \prod_{k=1}^{\infty} \frac{1 + \alpha^2 x^2/k^2\pi^2}{1 + x^2/k^2\pi^2}\ . \tag{5.15}$$

Each factor in this product is of the form

$$\frac{1+b^2x^2}{1+a^2x^2} = \frac{b^2}{a^2} + \frac{1-b^2/a^2}{1+a^2x^2}\ , \quad 0 \le b < a.$$

This shows that the function $\sinh \alpha x/\sinh x$ is positive definite for $0 \le \alpha \le 1$. In the same way using (5.14) one can see that $\cosh \alpha x/\cosh x$ is positive definite for $-1 \le \alpha \le 1$. The function

$$\frac{x \cosh \alpha\, x}{\sinh x} = \frac{x/2}{\sinh x/2}\ \frac{\cosh \alpha x}{\cosh x/2}$$

is positive definite for $-1/2 \le \alpha \le 1/2$, as it is the product of two such functions.

5.2.11

The integral

$$\frac{\tanh x}{x} = \int_0^1 \frac{\cosh \alpha x}{\cosh x} d\alpha$$

shows that $\tanh x/x$ is a positive definite function.

(Once again, it is natural to assign the functions $\sinh x/x$, $\sinh \alpha x/x$ and $\tanh x/x$ the values 1, α and 1, respectively, at $x = 0$. Then the functions under consideration are continuous and positive definite. Assigning them larger values at 0 destroys continuity but not positive definiteness.)

5.2.12

One more way of constructing positive definite functions is by convolutions of functions in $L_2(\mathbb{R})$. For any function f let $\tilde{f}$ be the function defined as $\tilde{f}(x) = \overline{f(-x)}$. If $f \in L_2(\mathbb{R})$ then the function $\varphi = f * \tilde{f}$ defined as

$$\varphi(x) = \int_{-\infty}^{\infty} f(x-t)\, \tilde{f}(t)\, dt$$

is a continuous function vanishing at ∞. It is a positive definite function since

$$\begin{aligned}\sum_{r,s=0}^{N-1} \varphi(x_r - x_s)\, \xi_r\, \overline{\xi}_s &= \sum_{r,s=0}^{N-1} \xi_r\, \overline{\xi}_s \int_{-\infty}^{\infty} f(x_r - x_s - t)\overline{f(-t)}\, dt \\ &= \sum_{r,s=0}^{N-1} \xi_r\, \overline{\xi}_s \int_{-\infty}^{\infty} f(x_r - t)\overline{f(x_s - t)}\, dt \\ &= \int_{-\infty}^{\infty} \left| \sum_{r=0}^{N-1} \xi_r\, f(x_r - t) \right|^2 dt \\ &\geq 0.\end{aligned}$$

5.2.13

Let R be a positive real number. The *tent function (with base R)* is defined as

$$\Delta_R(x) = \begin{cases} 1 - |x|/R & \text{if } |x| \leq R, \\ 0 & \text{if } |x| \geq R. \end{cases} \tag{5.16}$$

A calculation shows that

$$\Delta_R = \chi_{[-R/2,R/2]} \; * \; \chi_{[-R/2,R/2]}.$$

So, $\Delta_R(x)$ is positive definite for all $R > 0$.

5.2.14

For $R > 0$, let δ_R be the continuous function defined as

$$\delta_R(x) = \frac{R}{2\pi}\left(\frac{\sin\; Rx/2}{Rx/2}\right)^2. \tag{5.17}$$

From 5.2.6 it follows that δ_R is positive definite. The family $\{\delta_R\}_{R>0}$ is called the *Fejér kernel* on $\mathbb{R}$. It has the following properties (required of any "summability kernel" in Fourier analysis):

(i) $\delta_R(x) \geq 0$ for all x, and for all $R > 0$.

(ii) For every $a > 0$, $\delta_R(x) \to 0$ uniformly outside $[-a, a]$ as $R \to \infty$.

(iii) $\lim\limits_{R\to\infty} \int\limits_{|x|>a} \delta_R(x)dx = 0$ for every $a > 0$.

(iv) $\int\limits_{-\infty}^{\infty} \delta_R(x)dx = 1$ for all $R > 0$.

Property (iv) may be proved by contour integration, for example.

The functions Δ_R and δ_R are Fourier transforms of each other (up to constant factors). So the positive definiteness of one follows from the nonnegativity of the other.

5.2.15

In this section we consider functions like the tent functions of Section 5.2.13.

a Let φ be any continuous, nonnegative, even function. Suppose $\varphi(x) = 0$ for $|x| \geq R$, and φ is convex and monotonically decreasing in the interval $[0, R)$. Then φ is a uniform limit of sums of the form $\sum\limits_{k=1}^{n} a_k \triangle_{R_k}$, where $a_k \geq 0$ and $R_k \leq R$. It follows from 5.2.13 that φ is positive definite.

b The condition that φ is supported in $[-R, R]$ can be dropped. Let φ be any continuous, nonnegative, even function that is convex and monotonically decreasing in $[0, \infty)$. Let $b = \lim_{x\to\infty} \varphi(x)$. Then φ is a uniform limit of sums of the form $b + \sum_{k=1}^{n} a_k \triangle_{R_k}$, where $a_k \geq 0$ and $R_k > 0$. Hence φ is positive definite. This is *Pólya's Theorem.*

c Let φ be any function satisfying the conditions in Part **a** of this section, and extend it to all of $\mathbb{R}$ as a periodic function with period $2R$. Since φ is even, the Fourier expansion of φ does not contain any sin terms. It can be seen from the convexity of φ in $(0, R)$ that the coefficients a_n in the Fourier expansion

$$\varphi(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos \frac{n\pi}{R} x$$

are nonnegative. Hence φ is positive definite.

5.2.16

Using 5.2.15 one can see that the following functions are positive definite:

(i) $\varphi(x) = \frac{1}{1+|x|}$,

(ii) $\varphi(x) = \begin{cases} 1 - |x| & \text{for} \quad 0 \leq |x| \leq 1/2, \\ \frac{1}{4|x|} & \text{for} \quad |x| \geq 1/2, \end{cases}$

(iii) $\varphi(x) = \exp(-|x|^a)$, $0 \leq a \leq 1$.

The special case $a = 1$ of (iii) was seen in 5.2.8. The next theorem provides a further extension.

5.2.17 Theorem

The function $\varphi(x) = \exp(-|x|^a)$ *is positive definite for* $0 \leq a \leq 2$.

Proof. Let $0 < a < 2$. A calculation shows that

$$|x|^a = C_a \int_{-\infty}^{\infty} \frac{1 - \cos xt}{|t|^{1+a}} \, dt,$$

where

$$C_a = \frac{1}{2}\left[\int_0^\infty \frac{1-\cos t}{t^{1+a}}\, dt\right]^{-1}.$$

(The assumption $a < 2$ is needed to ensure that this last integral is convergent. At 0 the numerator in the integrand has a zero of order 2.) Thus we have

$$-|x|^a = \int_{-\infty}^{\infty} \frac{\cos xt - 1}{|t|^{1+a}}\, d\mu(t),$$

where $d\mu(t) = C_a\, dt$. Let φ_n be defined as

$$\begin{aligned}\varphi_n(x) &= \int_{|t|>1/n} \frac{\cos xt - 1}{|t|^{1+a}}\, d\mu(t)\\ &= \int_{|t|>1/n} \frac{\cos xt}{|t|^{1+a}}\, d\mu(t) - \int_{|t|>1/n} \frac{1}{|t|^{1+a}}\, d\mu(t).\end{aligned}$$

The second integral in the last line is a number, say c_n, while the first is a function, say $g_n(x)$. This function is positive definite since $\cos xt$ is positive definite for all t. So, for each n, the function $\exp \varphi_n(x)$ is positive definite by Exercise 5.1.5 (iii). Since $\lim_{n\to\infty} \exp \varphi_n(x) = \exp(-|x|^a)$, this function too is positive definite for $0 < a < 2$. Again, by continuity, this is true for $a = 2$ as well. ■

For $a > 2$ the function $\exp(-|x|^a)$ is *not* positive definite. This is shown in Exercise 5.5.8.

5.2.18

The equality

$$\begin{bmatrix} 1 & 1 & 1\\ 1 & 2 & 2\\ 1 & 2 & 3\end{bmatrix} = \begin{bmatrix} 1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix} + \begin{bmatrix} 0 & 0 & 0\\ 0 & 1 & 1\\ 0 & 1 & 1\end{bmatrix} + \begin{bmatrix} 0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 1\end{bmatrix}$$

shows that the matrix on the left-hand side is positive. Thus the $n \times n$ matrix A with entries $a_{ij} = \min(i,j)$ is positive. This can be used to see that the kernel $K(x, y) = \min(x, y)$ is positive definite on $[0, \infty) \times [0, \infty)$.

5.2.19 Exercise

Let B be the $n \times n$ matrix with entries $b_{ij} = 1/\max(i, j)$. Show that this matrix is positive by an argument similar to the one in 5.2.18.

Note that if A is the matrix in 5.2.18, then $B = DAD$, where $D = \operatorname{diag}(1, 1/2, \dots, 1/n)$.

5.2.20 Exercise

Let $\lambda_1, \dots, \lambda_n$ be positive real numbers. Let A and B be the $n \times n$ matrices whose entries are $a_{ij} = \min(\lambda_i, \lambda_j)$ and $b_{ij} = 1/\max(\lambda_i, \lambda_j)$, respectively. Show that A and B are positive definite.

5.2.21 Exercise

Show that the matrices A and B defined in Exercise 5.2.20 are infinitely divisible.

5.2.22 Exercise

Let $0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ and let A be the symmetric matrix whose entries a_{ij} are defined as $a_{ij} = \lambda_i/\lambda_j$ for $1 \le i \le j \le n$. Show that A is infinitely divisible.

5.2.23 Exercise

Let $\lambda_1, \lambda_2, \dots, \lambda_n$ be real numbers. Show that the matrix A whose entries are

$$a_{ij} = \frac{1}{1 + |\lambda_i - \lambda_j|}$$

is infinitely divisible. [Hint: Use Pólya's Theorem.]

5.3 LOEWNER MATRICES

In this section we resume, and expand upon, our discussion of operator monotone functions. Recall some of the notions introduced at the end of Chapter 2. Let $C^1(I)$ be the space of continuously differentiable real-valued functions on an open interval I. The *first divided difference* of a function f in $\mathrm{C}^1(I)$ is the function $f^{[1]}$ defined on $I \times I$ as

$$f^{[1]}(\lambda, \mu) = \frac{f(\lambda) - f(\mu)}{\lambda - \mu} \text{ if } \lambda \neq \mu,$$
$$f^{[1]}(\lambda, \lambda) = f'(\lambda).$$

Let $\mathbb{H}_n(I)$ be the collection of all $n \times n$ Hermitian matrices whose eigenvalues are in I. This is an open subset in the real vector space $\mathbb{H}_n$ consisting of all Hermitian matrices. The function f induces a map from $\mathbb{H}_n(I)$ into $\mathbb{H}_n$.

If $f \in \mathrm{C}^1(I)$ and $A \in \mathbb{H}_n(I)$ we define $f^{[1]}(A)$ as the matrix whose i, j entry is $f^{[1]}(\lambda_i, \lambda_j)$, where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A. This is called the *Loewner matrix* of f at A.

The function f on $\mathbb{H}_n(I)$ is differentiable. Its derivative at A, denoted as $Df(A)$, is a linear map on $\mathbb{H}_n$ characterized by the condition

$$||f(A+H) - f(A) - Df(A)(H)|| = o(||H||) \tag{5.18}$$

for all $H \in \mathbb{H}_n$. We have

$$Df(A)(H) = \left.\frac{d}{dt}\right|_{t=0} f(A+tH). \tag{5.19}$$

An interesting expression for this derivative in terms of Loewner matrices is given in the following theorem.

5.3.1 Theorem

Let $f \in \mathrm{C}^1(I)$ and $A \in \mathbb{H}_n(I)$. Then

$$Df(A)(H) = f^{[1]}(A) \circ H, \tag{5.20}$$

where $\circ$ denotes the Schur product in a basis in which A is diagonal.

One proof of this theorem can be found in MA (Theorem V.3.3). Here we give another proof based on different ideas.

Let $[X, Y]$ stand for the Lie bracket: $[X, Y] = XY - YX$. If X is Hermitian and Y skew-Hermitian, then $[X, Y]$ is Hermitian.

5.3.2 Theorem

Let $f \in \mathrm{C}^1(I)$ and $A \in \mathbb{H}_n(I)$. Then for every skew-Hermitian matrix K

$$Df(A)([A, K]) = [f(A), K]. \tag{5.21}$$

Proof. The exponential e^{tK} is a unitary matrix for all $t \in \mathbb{R}$. From the series representation of e^{tK} one can see that

$$\begin{aligned}[f(A), K] &= \frac{d}{dt}\bigg|_{t=0} e^{-tK} f(A) e^{tK} \\ &= \frac{d}{dt}\bigg|_{t=0} f(e^{-tK} A\, e^{tK}) \\ &= \frac{d}{dt}\bigg|_{t=0} f(A + t[A, K] + o(t)).\end{aligned}$$

Since f is in the class C^1, this is equal to

$$\frac{d}{dt}\bigg|_{t=0} f(A + t[A, K]) = Df(A)([A, K])\ . \qquad \blacksquare$$

For each $A \in \mathbb{H}_n$, the collection

$$\mathcal{C}_A = \{[A, K] : K^* = -K\}$$

is a linear subspace of $\mathbb{H}_n$. On $\mathbb{H}_n$ we have an inner product $\langle X, Y \rangle = \operatorname{tr} XY$. With respect to this inner product, the orthogonal complement of $\mathcal{C}_A$ is the space

$$\mathcal{Z}_A = \{H \in \mathbb{H}_n : [A, H] = O\}.$$

(It is easy to prove this. If H commutes with A, then

$$\langle H, [A, K] \rangle = \text{tr}\ H(AK - KA) = \text{tr}\ (HAK - HKA) = 0.)$$

Proof of Theorem 5.3.1. Choose an orthonormal basis in which $A = \text{diag}(\lambda_1, \ldots, \lambda_n)$. Let $H \in \mathcal{C}_A$; i.e., $H = [A, K]$ for some skew-Hermitian matrix K. By (5.21), $Df(A)(H) = [f(A), K]$. The entries of this matrix are

$$\begin{aligned}(f(\lambda_i) - f(\lambda_j))\, k_{ij} &= \frac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j}\,(\lambda_i - \lambda_j)\, k_{ij} \\ &= \frac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j}\, h_{ij}.\end{aligned}$$

These are the entries of $f^{[1]}(A) \circ H$ also. Thus the two sides of (5.20) are equal when $H \in \mathcal{C}_A$. Now let H belong to the complementary space $\mathcal{Z}_A$. The theorem will be proved if we show that the equality (5.20) holds in this case too. But this is easy. Since H commutes with A, we may assume H too is diagonal, $H = \text{diag}(h_1, \ldots, h_n)$. In this case the two sides of (5.20) are equal to the diagonal matrix with entries $f'(\lambda_i)h_i$ on the diagonal. ■

The next theorem says that f is operator monotone on I if and only if for all n and for all $A \in \mathbb{H}_n(I)$ the Loewner matrices $f^{[1]}(A)$ are positive. (This is a striking analogue of the statement that a real function f is monotonically increasing if and only if $f'(t) \geq 0$.)

5.3.3 Theorem

Let $f \in C^1(I)$. Then f is operator monotone on I if and only if $f^{[1]}(A)$ is positive for every Hermitian matrix A whose eigenvalues are contained in I.

Proof. Suppose f is operator monotone. Let $A \in \mathbb{H}_n(I)$ and let H be the positive matrix with all its entries equal to 1. For small positive t, $A + tH$ is in $\mathbb{H}_n(I)$. We have $A + tH \geq A$, and hence $f(A + tH) \geq f(A)$. This implies $Df(A)(H) \geq O$. For this H, the right-hand side of (5.20) is just $f^{[1]}(A)$, and we have shown this is positive.

To prove the converse, let $A_0,\ A_1$ be matrices in $\mathbb{H}_n(I)$ with $A_1 \geq A_0$. Let $A(t) = (1-t)A_0 + tA_1,\ 0 \leq t \leq 1$. Then $A(t)$ is in $\mathbb{H}_n(I)$. Our hypothesis says that $f^{[1]}(A(t))$ is positive. The derivative $A'(t) = A_1 - A_0$ is positive, and hence the Schur product $f^{[1]}(A(t)) \circ A'(t)$ is positive. By Theorem 5.3.1 this product is equal to $Df(A(t))(A'(t))$. Since

$$f(A_1) - f(A_0) = \int_0^1 Df(A(t))(A'(t))dt$$

and the integrand is positive for all t, we have $f(A_1) \geq f(A_0)$. ■

We have seen some examples of operator monotone functions in Section 4.2. Theorem 5.3.3 provides a direct way of proving operator monotonicity of these and other functions. The positivity of the Loewner matrices $f^{[1]}(A)$ is proved by associating with them some positive definite functions. Some examples follow.

5.3.4

The function

$$f(t) = \frac{at+b}{ct+d}\ ,\ a,b,c,d \in \mathbb{R},\ ad-bc > 0$$

is operator monotone on any interval I that does not contain the point $-d/c$.

To see this write down the Loewner matrix $f^{[1]}(A)$ for any $A \in \mathbb{H}_n(I)$. If $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A, this Loewner matrix has entries

$$\frac{ad-bc}{(c\lambda_i + d)(c\lambda_j + d)}.$$

This matrix is congruent to the matrix with all entries 1, and is therefore positive.

5.3.5

The function $f(t) = t^r$ is operator monotone on $(0, \infty)$ for $0 \leq r \leq 1$.

A Loewner matrix for this function is a matrix V with entries

$$v_{ij} = \frac{\lambda_i^r - \lambda_j^r}{\lambda_i - \lambda_j}, \quad i \neq j,$$
$$v_{ii} = r\, \lambda_i^{r-1} \quad \text{for all } i.$$

The numbers λ_i are positive and can, therefore, be written as e^{x_i} for some x_i. We have then

$$\begin{aligned} v_{ij} &= \frac{e^{rx_i} - e^{rx_j}}{e^{x_i} - e^{x_j}} \\ &= \frac{e^{rx_i/2}}{e^{x_i/2}} \, \frac{e^{r(x_i - x_j)/2} - e^{r(x_j - x_i)/2}}{e^{(x_i - x_j)/2} - e^{(x_j - x_i)/2}} \, \frac{e^{rx_j/2}}{e^{x_j/2}} \\ &= \frac{e^{rx_i/2}}{e^{x_i/2}} \, \frac{\sinh r(x_i - x_j)/2}{\sinh\,(x_i - x_j)/2} \, \frac{e^{rx_j/2}}{e^{x_j/2}}. \end{aligned}$$

This matrix is congruent to the matrix with entries

$$\frac{\sinh r(x_i - x_j)/2}{\sinh\,(x_i - x_j)/2}.$$

Since $\sinh rx/(\sinh x)$ is a positive definite function for $0 \leq r \leq 1$ (see 5.2.10), this matrix is positive.

5.3.6 Exercise

The function $f(t) = t^r$ is not operator monotone on $(0, \infty)$ for any real number r outside $[0, 1]$.

5.3.7

The function $f(t) = \log t$ is operator monotone on $(0, \infty)$.

A Loewner matrix in this case has entries

$$v_{ij} = \frac{\log \lambda_i - \log \lambda_j}{\lambda_i - \lambda_j}, \quad i \neq j,$$
$$v_{ii} = \frac{1}{\lambda_i} \quad \text{for all } i.$$

The substitution $\lambda_i = e^{x_i}$ reduces this to

$$v_{ij} = \frac{x_i - x_j}{e^{x_i} - e^{x_j}} = \frac{1}{e^{x_i/2}} \frac{(x_i - x_j)/2}{\sinh\,(x_i - x_j)/2} \frac{1}{e^{x_j/2}}.$$

This matrix is positive since the function $x/(\sinh x)$ is positive definite. (See 5.2.9.)

Another proof of this is obtained from the equality

$$\frac{\log \lambda_i - \log \lambda_j}{\lambda_i - \lambda_j} = \int_0^\infty \frac{1}{(\lambda_i + t)(\lambda_j + t)}\, dt.$$

For each t the matrix $[[1/(\lambda_i + t)(\lambda_j + t)]]$ is positive. (One more proof of operator monotonicity of the log function was given in Exercise 4.2.5.)

5.3.8

The function $f(t) = \tan t$ is operator monotone on $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$.

In this case a Loewner matrix has entries

$$\begin{aligned} v_{ij} &= \frac{\tan \lambda_i - \tan \lambda_j}{\lambda_i - \lambda_j} \\ &= \frac{1}{\cos\,\lambda_i} \frac{\sin(\lambda_\mathrm{i} - \lambda_\mathrm{j})}{\lambda_i - \lambda_j} \frac{1}{\cos \lambda_j}\,. \end{aligned}$$

This matrix is positive since the function $\sin x/x$ is positive definite. (See 5.2.6.)

5.3.9 Exercise

For $0 \leq r \leq 1$ let f be the map $f(A) = A^r$ on the space of positive definite matrices. Show that

$$\|Df(A)\| = r\|A\|^{r-1}. \tag{5.22}$$

5.4 NORM INEQUALITIES FOR MEANS

The theme of this section is inequalities for norms of some expressions involving positive matrices. In the case of numbers they reduce to some of the most fundamental inequalities of analysis.

As a prototype consider the arithmetic-geometric mean inequality $\sqrt{ab} \le \frac{1}{2}(a+b)$ for positive numbers a, b. There are many different directions in which one could look for a generalization of this to positive matrices A, B. One version that involves the somewhat subtle concept of a matrix geometric mean is given in Section 4.1. Instead of matrices we could compare *numbers* associated with them. Thus, for example, we may ask whether

$$|||A^{1/2}B^{1/2}||| \le \frac{1}{2}|||A+B||| \tag{5.23}$$

for every unitarily invariant norm. This is indeed true. There is a more general version of this inequality that is easier to prove: we have

$$|||A^{1/2}XB^{1/2}||| \le \frac{1}{2}|||AX+XB||| \tag{5.24}$$

for every X. What makes it easier is a lovely trick. It is enough to prove (5.24) in the special case $A = B$. (The inequality (5.23) is a vacuous statement in this case.) Suppose we have proved

$$|||A^{1/2}XA^{1/2}||| \le \frac{1}{2}|||AX+XA||| \tag{5.25}$$

for all matrices X and positive A. Then given X and positive A, B we may replace A and X in (5.25) by the 2×2 block matrices $\begin{bmatrix} A & O \\ O & B \end{bmatrix}$ and $\begin{bmatrix} O & X \\ O & O \end{bmatrix}$. This gives the inequality (5.24).

Since the norms involved are unitarily invariant we may assume that A is diagonal, $A = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Then we have

$$A^{1/2}XA^{1/2} = Y \circ \left(\frac{AX+XA}{2}\right) \tag{5.26}$$

where Y is the matrix with entries

$$y_{ij} = \frac{2\sqrt{\lambda_i \lambda_j}}{\lambda_i + \lambda_j}\,. \tag{5.27}$$

This matrix is congruent to the Cauchy matrix—the one whose entries are $1/(\lambda_i + \lambda_j)$. Since that matrix is positive (Exercise 1.1.2) so is Y. All the diagonal entries of Y are equal to 1. So, using Exercise 2.7.12 we get the inequality (5.25) from (5.26).

The inequalities that follow are proved using similar arguments. Matrices that occur in the place of (5.27) are more complicated and their positivity is not as easy to establish. But in Section 5.2 we have done most of the work that is needed.

5.4.1 Theorem

Let A, B be positive and let X be any matrix. Then for $0 \le \nu \le 1$ we have

$$2|||A^{1/2}XB^{1/2}||| \le |||A^{\nu}XB^{1-\nu} + A^{1-\nu}XB^{\nu}||| \le |||AX + XB|||. \tag{5.28}$$

Proof. Follow the arguments used above in proving (5.24). To prove the second inequality in (5.28) we need to prove that the matrix Y whose entries are

$$y_{ij} = \frac{\lambda_i^{\nu}\lambda_j^{1-\nu} + \lambda_i^{1-\nu}\lambda_j^{\nu}}{\lambda_i + \lambda_j} \tag{5.29}$$

is positive for $0 < \nu < 1$. (When $\nu = 1/2$ this reduces to (5.27).) Writing

$$y_{ij} = \lambda_i^{1-\nu}\left(\frac{\lambda_i^{2\nu-1} + \lambda_j^{2\nu-1}}{\lambda_i + \lambda_j}\right)\lambda_j^{1-\nu}$$

we see that Y is congruent to the matrix Z whose entries are

$$z_{ij} = \frac{\lambda_i^\alpha + \lambda_j^\alpha}{\lambda_i + \lambda_j}\ , \text{ where } -1 < \alpha < 1.$$

This matrix is like the one in 5.3.5. The argument used there reduces the question of positivity of Z to that of positive definiteness of the function $\cosh \alpha x/(\cosh\, x)$ for $-1 < \alpha < 1$. In 5.2.10 we have seen that this function is indeed positive definite. The proof of the first inequality in (5.28) is very similar to this, and is left to the reader. ■

5.4.2 Exercise

Show that for $0 \le \nu \le 1$

$$|||A^\nu X B^{1-\nu} - A^{1-\nu} X B^\nu||| \le |2\nu - 1|\ |||AX - XB|||. \tag{5.30}$$

5.4.3 Exercise

For the Hilbert-Schmidt norm we have

$$||A^\nu X A^{1-\nu}||_2 \le ||\nu AX + (1-\nu)XA||_2 \tag{5.31}$$

for positive matrices A and $0 < \nu < 1$. This is not always true for the operator norm $||\cdot||$.

5.4.4 Exercise

For any matrix Z let

$$\text{Re}\, Z = \frac{1}{2}\,(Z + Z^*)\ ,\ \ \text{Im}\, Z = \frac{1}{2i}\,(Z - Z^*).$$

Let A be a positive matrix and let X be a Hermitian matrix. Let $S = A^\nu X A^{1-\nu}$, $T = \nu AX + (1-\nu)XA$. Show that for $0 \le \nu \le 1$

$$|||\text{Re}\, S||| \le |||\text{Re}\, T|||\ ,\ \ |||\text{Im}\, S||| \le |||\text{Im}\, T|||.$$

In Chapter 4 we defined the logarithmic mean of a and b. This is the quantity

$$\begin{aligned}\frac{a-b}{\log a - \log b} &= \int_0^1 a^t\, b^{1-t}\, dt \\ &= \left[\int_0^1 \frac{dt}{ta+(1-t)b}\right]^{-1} = \left[\int_0^\infty \frac{dt}{(t+a)(t+b)}\right]^{-1}.\end{aligned} \tag{5.32}$$

A proof of the inequality (4.3) using the ideas of Section 5.3 is given below.

5.4.5 Lemma

$$\sqrt{ab} \leq \frac{a-b}{\log a - \log b} \leq \frac{a+b}{2}. \tag{5.33}$$

Proof. Put $a = e^x$ and $b = e^y$. A small calculation reduces the job of proving the first inequality in (5.33) to showing that $t \leq \sinh t$ for $t > 0$, and the second to showing that $\tanh t \leq t$ for all $t > 0$. Both these inequalities can be proved very easily. ■

5.4.6 Exercise

Show that for A, B positive and for every X

$$||A^{1/2}XB^{1/2}||_2 \leq \left|\left|\int_0^1 A^t X B^{1-t} dt\right|\right|_2 \leq \frac{1}{2}\, ||AX + XB||_2\,. \tag{5.34}$$

This matrix version of the arithmetic-logarithmic-geometric mean inequality can be generalized to all unitarily invariant norms.

5.4.7 Theorem

For every unitarily invariant norm we have

$$\left|\left|\left|A^{1/2}XB^{1/2}\right|\right|\right| \leq \left|\left|\left|\int_0^1 A^t X B^{1-t} dt\right|\right|\right| \leq \frac{1}{2}\, |||AX + XB|||. \tag{5.35}$$

Proof. The idea of the proof is very similar to that of Theorem 5.4.1. Assume $B = A$, and suppose A is diagonal with entries $\lambda_1, \ldots, \lambda_n$ on the diagonal. The matrix $A^{1/2}XA^{1/2}$ is obtained from $\int_0^1 A^t X A^{1-t} dt$ by entrywise multiplication with the matrix Y whose entries are

$$y_{ij} = \frac{\lambda_i^{1/2}\,\lambda_j^{1/2}(\log \lambda_i - \log \lambda_j)}{\lambda_i - \lambda_j}\,.$$

This matrix is congruent to one with entries

$$z_{ij} = \frac{\log \lambda_i - \log \lambda_j}{\lambda_i - \lambda_j}\,.$$

We have seen in 5.3.7 that this matrix is positive. That proves the first inequality in (5.35).

The matrix $\int_0^1 A^t X A^{1-t} dt$ is the Schur product of $\frac{1}{2}(AX + XA)$ with the matrix W whose entries are

$$w_{ij} = \frac{2(\lambda_i - \lambda_j)}{(\log \lambda_i - \log \lambda_j)(\lambda_i + \lambda_j)}\,.$$

Making the substitution $\lambda_i = e^{x_i}$, we have

$$w_{ij} = \frac{\tanh\,(x_i - x_j)/2}{(x_i - x_j)/2}\,.$$

This matrix is positive since the function $\tanh x/x$ is positive definite. (See 5.2.11.) That proves the second inequality in (5.35). ■

5.4.8 Exercise

A refinement of the inequalities (5.28) and (5.35) is provided by the assertion

$$\frac{1}{2}|||A^\nu X B^{1-\nu} + A^{1-\nu} X B^\nu||| \le |||\int_0^1 A^t X B^{1-t} dt|||$$

for $1/4 \le \nu \le 3/4$. Prove this using the fact that $(x \cosh \alpha x)/\sinh x$ is a positive definite function for $-1/2 \le \alpha \le 1/2$. (See 5.2.10.)

5.4.9 Exercise

Let H, K be Hermitian, and let X be any matrix. Show that

$$|||(\sin H)X(\cos K) \pm (\cos H)X(\sin K)||| \le |||HX \pm XK|||.$$

This is a matrix version of the inequality $|\sin x| \le |x|$.

5.4.10 Exercise

Let H, K and X be as above. Show that

$$|||HX \pm XK||| \le |||(\sinh H)X(\cosh K) \pm (\cosh H)X(\sinh K)|||.$$

5.4.11 Exercise

Let A, B be positive matrices. Show that

$$|||(\log A)X - X(\log B)||| \le |||A^{1/2}XB^{-1/2} - A^{-1/2}XB^{1/2}|||.$$

Hence, if H, K are Hermitian, then

$$|||HX - XK||| \le |||e^{H/2}Xe^{-K/2} - e^{-H/2}Xe^{K/2}|||$$

for every matrix X.

5.5 THEOREMS OF HERGLOTZ AND BOCHNER

These two theorems give complete characterizations of positive definite sequences and positive definite functions, respectively. They have important applications throughout analysis. For the sake of completeness we include proofs of these theorems here. Some basic facts from functional analysis and Fourier analysis are needed for the proofs. The reader is briefly reminded of these facts.

Let M[0, 1] be the space of complex finite Borel measures on the interval [0, 1]. This is equipped with a norm $||\mu|| = \int |d\mu|$, and is the Banach space dual of the space C[0, 1]. If $\int f d\mu_n$ converges to $\int f d\mu$ for every $f \in$ C[0, 1], we say that the sequence $\{\mu_n\}$ in M[0, 1] converges to μ in the weak* topology.

A basic fact about this convergence is the following theorem called *Helly's Selection Principle.*

5.5.1 Theorem

Let $\{\mu_n\}$ be a sequence of probability measures on [0, 1]. Then there exists a probability measure μ and a subsequence $\{\mu_m\}$ of $\{\mu_n\}$ such that μ_m converges in the weak topology to μ.*

Proof. The space C[0, 1] is a separable Banach space. Choose a sequence $\{f_j\}$ in C[0, 1] that includes the function **1** and whose linear combinations are dense in C[0, 1]. Since $||\mu_n|| = 1$, for each j we have $|\int f_j \, d\mu_n| \leq ||f_j||$ for all n. Thus for each j, $\{|\int f_j \, d\mu_n|\}$ is a bounded sequence of positive numbers. By the diagonal procedure, we can extract a subsequence $\{\mu_m\}$ such that for each j, the sequence $\int f_j \, d\mu_m$ converges to a limit, say ξ_j, as $m \to \infty$.

If $f = \sum a_j f_j$ is any (finite) linear combination of the f_j, let

$$\Lambda_0(f) := \sum a_j \xi_j = \lim_{m \to \infty} \int f \, d\mu_m.$$

This is a linear functional on the linear span of $\{f_j\}$, and $|\Lambda_0(f)| \leq ||f||$ for every f in this span. By continuity Λ_0 has an extension Λ to C[0, 1] that satisfies $|\Lambda(f)| \leq ||f||$ for all f in C[0, 1]. This linear functional Λ is positive and unital. So, by the Riesz Representation Theorem, there exists a probability measure μ on [0, 1] such that $\Lambda(f) = \int f \, d\mu$ for all $f \in$ C[0, 1].

Finally, we know that $\int f \, d\mu_m$ converges to $\int f \, d\mu$ for every f in the span of $\{f_j\}$. Since such f are dense and the μ_m are uniformly bounded, this convergence persists for every f in C[0, 1]. ■

Theorem 5.5.1 is also a corollary of the Banach Alaoglu theorem. This says that the closed unit ball in the dual space of a Banach space is compact in the weak* topology. If a Banach space X is separable, then the weak* topology on the closed unit ball of its dual X^* is metrizable.

5.5.2 Herglotz' Theorem

Let $\{a_n\}_{n\in\mathbb{Z}}$ be a positive definite sequence and suppose $a_0 = 1$. Then there exists a probability measure μ on $[-\pi, \pi]$ such that

$$a_n = \int_{-\pi}^{\pi} e^{-inx} d\mu(x). \tag{5.36}$$

Proof. The positive definiteness of $\{a_n\}$ implies that for every real x we have

$$\sum_{r,s=0}^{N-1} a_{r-s} \; e^{i(r-s)x} \geq 0.$$

This inequality can be expressed in another form

$$N \sum_{k=-(N-1)}^{N-1} \left(1 - \frac{|k|}{N}\right) a_k e^{ikx} \geq 0.$$

Let $f_N(x)$ be the function given by the last sum. Then

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} f_N(x) dx = a_0 = 1.$$

For any Borel set E in $[-\pi, \pi]$, let

$$\mu_N(E) = \frac{1}{2\pi} \int_E f_N(x) dx.$$

Then μ_N is a probability measure on $[-\pi, \pi]$. Apply Helly's selection principle to the sequence $\{\mu_N\}$. There exists a probability measure μ to which (a subsequence of) μ_N converges in the weak* topology. Thus for every n

$$\int_{-\pi}^{\pi} e^{-inx} d\mu(x) = \lim_{N\to\infty} \int_{-\pi}^{\pi} e^{-inx} d\mu_N(x)$$

$$= \lim_{N\to\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-inx} f_N(x)dx$$

$$= \lim_{N\to\infty} \left(1 - \frac{|n|}{N}\right) a_n$$

$$= a_n. \blacksquare$$

We remark that the sum

$$\sum_{k=-(N-1)}^{N-1} \left(1 - \frac{|k|}{N}\right) e^{ikx}$$

is called the Fejér kernel and is much used in the study of Fourier series.

The condition $a_0 = 1$ in the statement of Herglotz' theorem is an inessential normalization. This can be dropped; then μ is a finite positive measure with $||\mu|| = a_0$.

Bochner's theorem, in the same spirit as Herglotz', says that every continuous positive definite function on $\mathbb{R}$ is the Fourier-Stieltjes transform of a finite positive measure on $\mathbb{R}$. The proof needs some approximation arguments. For the convenience of the reader let us recall some basic facts.

For $f \in L_1(\mathbb{R})$ we write $\hat{f}$ for its Fourier transform defined as

$$\hat{f}(x) = \int_{-\infty}^{\infty} e^{-itx} f(t)dt.$$

This function is in $C_0(\mathbb{R})$, the class of continuous functions vanishing at ∞. We write

$$\check{f}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx} f(t)dt$$

for the inverse Fourier transform of f. If the function $\hat{f}$ is in $L_1(\mathbb{R})$ (and this is not always the case) then $f = (\hat{f})\check{}$.

The Fourier transform on the space $L_2(\mathbb{R})$ is defined as follows. Let $f \in L_2(\mathbb{R}) \cap L_1(\mathbb{R})$. Then $\hat{f}$ is defined as above. One can see that $\hat{f} \in L_2(\mathbb{R})$ and the map $f \mapsto (2\pi)^{-1/2}\hat{f}$ is an L_2-isometry on the space $L_2(\mathbb{R}) \cap L_1(\mathbb{R})$. This space is dense in $L_2(\mathbb{R})$. So the isometry defined on it has a unique extension to all of $L_2(\mathbb{R})$. This unitary operator on $L_2(\mathbb{R})$ is denoted again by $(2\pi)^{-1/2}\hat{f}$. The inverse of the map $f \mapsto \hat{f}$ is defined by inverting this unitary operator. The fact that the Fourier transform is a bijective map of $L_2(\mathbb{R})$ onto itself makes some operations in this space simpler.

Let δ_R be the function defined in 5.2.14. The family $\{\delta_N\}$ is an *approximate identity*: as $N \to \infty$, the convolution $\delta_N * g$ converges to g in an appropriate sense. The "appropriate sense" for us is the following.

If g is either an element of $L_1(\mathbb{R})$, or a bounded measurable function, then

$$\lim_{N\to\infty} (\delta_N * g)(x) := \lim_{N\to\infty} \int_{-\infty}^{\infty} \delta_N(x-t)g(t)dt = g(x) \ \text{ a.e.} \tag{5.37}$$

In the discussion that follows we ignore constant factors involving 2π. These do not affect our conclusions in any way.

The Fourier transform "converts convolution into multiplication;" i.e.,

$$\widehat{f * g} = \hat{f}\hat{g} \quad \text{for all } f, g \in L_1(\mathbb{R}).$$

The Riesz representation theorem and Helly's selection principle have generalizations to the real line. The space $\mathrm{C}_0(\mathbb{R})$ is a separable Banach space. Its dual is the space $\mathrm{M}(\mathbb{R})$ of finite Borel measures on $\mathbb{R}$ with norm $||\mu|| = \int |d\mu|$. Every bounded sequence $\{\mu_n\}$ in $\mathrm{M}(\mathbb{R})$ has a weak* convergent subsequence $\{\mu_m\}$; i.e., for every $f \in \mathrm{C}_0(\mathbb{R})$, $\int f d\mu_m$ converges to $\int f d\mu$ as $m \to \infty$. This too is a special case of the Banach-Alaoglu theorem.

5.5.3 Bochner's Theorem

Let φ be any function on the real line that is positive definite and continuous at 0. Then there exists a finite positive measure μ such that

$$\varphi(x) = \int_{-\infty}^{\infty} e^{-itx} d\mu(t). \tag{5.38}$$

Proof. By Lemma 5.1.2, φ is continuous everywhere. Suppose in addition that $\varphi \in L_1(\mathbb{R})$. Using (5.6) we see that

$$\int_{-\infty}^{\infty} |\varphi(x)|^2 dx \leq \varphi(0) \int_{-\infty}^{\infty} |\varphi(x)| dx.$$

Thus φ is in the space $L_2(\mathbb{R})$ also. Hence, there exists $f \in L_2(\mathbb{R})$ such that

$$f(t) = \check{\varphi}(t) = \int_{-\infty}^{\infty} e^{itx} \varphi(x) dx. \tag{5.39}$$

Let $\Delta_N(x)$ be the tent function defined in (5.16). Then

$$\int_{-\infty}^{\infty} e^{itx} \varphi(x) \Delta_N(x) dx = \int_{-N}^{N} e^{itx} \varphi(x) \left(1 - \frac{|x|}{N}\right) dx. \tag{5.40}$$

This integral (of a continuous function over a bounded interval) is a limit of Riemann sums. Let $x_j = j\,N/K$, $-K \leq j \leq K$. The last integral is the limit, as $K \to \infty$, of sums

$$\sum_{j=-(K-1)}^{K-1} e^{itx_j} \varphi(x_j) \left(1 - \frac{|x_j|}{N}\right) \frac{N}{K}.$$

These sums can be expressed in another way:

$$c(K, N) \sum_{r,s=0}^{K-1} e^{it(x_r - x_s)} \varphi(x_r - x_s) \tag{5.41}$$

where $c(K, N)$ is a positive number. (See the proof of Herglotz' theorem where two sums of this type were involved.) Since φ is positive definite, the sum in (5.41) is nonnegative. Hence, the integral (5.41), being the limit of such sums, is nonnegative. As $N \to \infty$ the integral in (5.40) tends to the one in (5.39). So, that integral is nonnegative too. Thus $f(t) \geq 0$.

Now let φ be any continuous positive definite function and let

$$\varphi_n(x) = e^{-x^2/n}\varphi(x).$$

Since φ is bounded, φ_n is integrable. Since $\varphi(x)$ and $e^{-x^2/n}$ are positive definite, so is their product $\varphi_n(x)$. Thus by what we have proved in the preceding paragraph, for each n

$$\varphi_n = \hat{f}_n, \quad \text{where} \quad f_n \in L_2(\mathbb{R}) \quad \text{and} \quad f_n \geq 0 \quad \text{a.e.}$$

We have the relation $\delta_N * \varphi_n = (\Delta_N f_n)\hat{}$, i.e.,

$$\int_{-\infty}^{\infty} \delta_N(x-t)\varphi_n(t)dt = \int_{-\infty}^{\infty} e^{-itx}\Delta_N(t)f_n(t)dt. \tag{5.42}$$

At $x = 0$ this gives

$$\begin{aligned}\int_{-\infty}^{\infty} \Delta_N(t)f_n(t)dt &= \int_{-\infty}^{\infty} \delta_N(-t)\varphi_n(t)dt \\ &\leq \varphi_n(0)\int_{-\infty}^{\infty} \delta_N(-t)dt \\ &= \varphi(0).\end{aligned}$$

Let $N \to \infty$. This shows

$$\int_{-\infty}^{\infty} f_n(t)dt \leq \varphi(0) \quad \text{for all } n,$$

i.e., $f_n \in L_1(\mathbb{R})$ and $||f_n||_1 \leq \varphi(0)$. Let $d\mu_n(t) = f_n(t)dt$. Then $\{\mu_n\}$ are positive measures on $\mathbb{R}$ and $||\mu_n|| \leq \varphi(0)$. So, by Helly's selection principle, there exists a positive measure μ, with $||\mu|| \leq \varphi(0)$, to which (a subsequence of) μ_n converges in the weak* topology.

The equation (5.42) says

$$\int_{-\infty}^{\infty} \delta_N(x-t)\varphi_n(t)dt = \int_{-\infty}^{\infty} e^{-itx}\Delta_N(t)d\mu_n(t). \tag{5.43}$$

Keep N fixed and let $n \to \infty$. For the right-hand side of (5.43) use the weak* convergence of μ_n to μ, and for the left-hand side the Lebesgue-dominated convergence theorem. This gives

$$\int_{-\infty}^{\infty} \delta_N(x-t)\varphi(t)dt = \int_{-\infty}^{\infty} e^{-itx}\Delta_N(t)d\mu(t). \tag{5.44}$$

Now let $N \to \infty$. Since φ is a bounded measurable function, by (5.37) the left-hand side of (5.44) goes to $\varphi(x)$ a.e. The right-hand side converges by the bounded convergence theorem. This shows

$$\varphi(x) = \int_{-\infty}^{\infty} e^{-itx} d\mu(t) \quad \text{a.e.}$$

Since the two sides are continuous functions of x, this equality holds everywhere. ■

Of the several examples of positive definite functions in Section 5.2 some were shown to be Fourier transforms of nonnegative integrable functions. (See 5.2.5 - 5.2.8.) One can do this for some of the other functions too.

5.5.4

The list below gives some functions φ and their Fourier transforms $\hat{\varphi}$ (ignoring constant factors).

(i) $\varphi(x) = \dfrac{x}{\sinh x}, \quad \hat{\varphi}(t) = \dfrac{1}{\cosh^2(t\pi/2)}$.

(ii) $\varphi(x) = \dfrac{1}{\cosh x}, \quad \hat{\varphi}(t) = \dfrac{1}{\cosh\,(t\pi/2)}$.

(iii) $\varphi(x) = \dfrac{\sinh \alpha x}{\sinh x}, \quad \hat{\varphi}(t) = \dfrac{\sin \alpha\pi}{\cosh t\pi + \cos \alpha\pi}, \quad 0 < \alpha < 1.$

(iv) $\varphi(x) = \dfrac{\cosh \alpha x}{\cosh x}, \quad \hat{\varphi}(t) = \dfrac{\cos \alpha\pi/2 \cosh t\pi/2}{\cosh t\pi + \cos \alpha\pi}, \quad -1 < \alpha < 1.$

(v) $\varphi(x) = \dfrac{\tanh x}{x}, \quad \hat{\varphi}(t) = \log \coth \dfrac{\pi t}{4}, \quad t > 0.$

Let φ be a continuous positive definite function. Then the measure μ associated with φ via the formula (5.38) is a probability measure if and only if $\varphi(0) = 1$. In this case φ is called a *characteristic function.*

5.5.5 Proposition

Let φ be a characteristic function. Then

$$1 - \operatorname{Re} \varphi(2^n x) \le 4^n (1 - \operatorname{Re} \varphi(x)),$$

for all x and $n = 1, 2, \ldots$.

Proof. By elementary trigonometry

$$1 - \cos tx = 2 \sin^2 \frac{tx}{2} \ge 2 \sin^2 \frac{tx}{2} \cos^2 \frac{tx}{2} = \frac{1}{2} \sin^2 tx = \frac{1}{4}(1 - \cos 2tx).$$

An iteration leads to the inequality

$$1 - \cos tx \ge \frac{1}{4^n}(1 - \cos 2^n tx).$$

From (5.38) we have

$$\begin{aligned} 1 - \operatorname{Re} \varphi(x) &= \int_{-\infty}^{\infty} (1 - \cos tx) d\mu(t) \\ &\ge \frac{1}{4^n} \int_{-\infty}^{\infty} (1 - \cos 2^n tx) d\mu(t) \\ &= \frac{1}{4^n} [1 - \operatorname{Re} \varphi(2^n x)]. \quad \blacksquare \end{aligned}$$

5.5.6 Corollary

Suppose φ is a positive definite function and $\varphi(x) = \varphi(0) + o(x^2)$; i.e.,

$$\lim_{x\to 0} \frac{\varphi(x) - \varphi(0)}{x^2} = 0.$$

Then φ is a constant.

Proof. We may assume that $\varphi(0) = 1$. Then using the Proposition above we have for all x and n

$$1 - \operatorname{Re}\varphi(x) \leq 4^n[1 - \operatorname{Re}\varphi(x/2^n)] = \frac{1 - \operatorname{Re}\varphi(x/2^n)}{(x/2^n)^2}\, x^2.$$

The hypothesis on φ implies that the last expression goes to zero as $n \to \infty$. Hence, $\operatorname{Re}\varphi(x) = 1$ for all x. But then $\varphi(x) \equiv 1$. ∎

5.5.7 Exercise

Suppose φ is a characteristic function, and $\varphi(x) = 1 + o(x) + o(x^2)$ in a neighbourhood of 0, where $o(x)$ is an odd function. Then $\varphi \equiv 1$. [Hint: consider $\varphi(x)\varphi(-x)$.]

5.5.8 Exercise

The functions e^{-x^4}, $1/(1+x^4)$, and $e^{-|x|^a}$ for $a > 2$, are not positive definite.

Bochner's theorem can be used also to show that a certain function is *not* positive definite by showing that its Fourier transform is not everywhere nonnegative.

5.5.9 Exercise

Use the method of residues to show that for all $t > 0$

$$\int_{-\infty}^{\infty} \frac{\cos(tx)}{1+x^4}\, dx = \frac{\pi}{\sqrt{2}}\, e^{-t/\sqrt{2}} \left(\cos\frac{t}{\sqrt{2}} + \sin\frac{t}{\sqrt{2}}\right).$$

It follows from Bochner's theorem that the function $f(x) = 1/(1+x^4)$ is not positive definite.

5.6 SUPPLEMENTARY RESULTS AND EXERCISES

5.6.1 Exercise

Let U be a unitary operator on any separable Hilbert space $\mathcal{H}$. Show that for each unit vector x in $\mathcal{H}$ the sequence

$$a_n = \langle x, U^n x \rangle \tag{5.45}$$

is positive definite.

This observation is the first step on one of the several routes to the spectral theorem for operators in Hilbert space. We indicate this briefly.

Let U be a unitary operator on $\mathcal{H}$. By Exercise 5.6.1 and Herglotz' theorem, for each unit vector x in $\mathcal{H}$, there exists a probability measure μ_x on the interval $[-\pi, \pi]$ such that

$$\langle x, U^n x \rangle = \int_{-\pi}^{\pi} e^{int} d\mu_x(t). \tag{5.46}$$

Using a standard technique called polarisation, one can obtain from this, for each pair x, y of unit vectors a complex measure $\mu_{x,y}$ such that

$$\langle y, U^n x \rangle = \int_{-\pi}^{\pi} e^{int} d\mu_{x,y}(t). \tag{5.47}$$

Now for each Borel subset $E \subset [-\pi, \pi]$ let $P(E)$ be the operator on $\mathcal{H}$ defined by the relation

$$\langle y, P(E)x \rangle = \mu_{x,y}(E) \quad \text{for all } x, y. \tag{5.48}$$

It can be seen that $P(E)$ is an orthogonal projection and that $P(\cdot)$ is countably additive on the Borel σ-algebra of $[-\pi, \pi]$. In other words $P(\cdot)$ is a *projection-valued measure.* We can then express U as an integral

$$U = \int_{-\pi}^{\pi} e^{it} dP(t). \tag{5.49}$$

This is the spectral theorem for unitary operators. The spectral theorem for self-adjoint operators can be obtained from this using the Cayley transform.

5.6.2 Exercise

Let B be an $n \times n$ Hermitian matrix. Show that for each unit vector u the function

$$\varphi(t) = \langle u, e^{itB} u \rangle$$

is a positive definite function on $\mathbb{R}$. Use this to show that the functions $\operatorname{tr} e^{itB}$ and $\det e^{itB}$ are positive definite.

5.6.3 Exercise

Let A, B be $n \times n$ Hermitian matrices and let

$$\varphi(t) = \operatorname{tr} e^{A+itB}. \tag{5.50}$$

Is φ a positive definite function? Show that this is so if A and B commute.

The general case of the question raised above is a long-standing open problem in quantum statistical mechanics. The Bessis-Moussa-Villani conjecture says that the function φ in (5.50) is positive definite for all Hermitian matrices A and B.

The purpose of the next three exercises is to calculate Fourier transforms of some functions that arose in our discussion.

5.6.4 Exercise

Let $\varphi(x) = 1/\cosh x$. Its Fourier transform is

$$\widehat{\varphi}(t) = \int_{-\infty}^{\infty} \frac{e^{-itx}}{\cosh x} dx.$$

This integral may be evaluated by the method of residues. Let f be the function

$$f(z) = \frac{e^{-itz}}{\cosh z}.$$

Then

$$f(z + i\pi) = -e^{t\pi} f(z) \text{ for all } z.$$

For any $R > 0$ the rectangular contour with vertices $-R, R, R+i\pi$ and $-R + i\pi$ contains one simple pole, $z = i\pi/2$, of f inside it. Integrate

f along this contour and then let $|R| \to \infty$. The contribution of the two vertical sides goes to zero. So

$$\int_{-\infty}^{\infty} \frac{e^{-itx}}{\cosh x} dx = \frac{2\pi i}{1+e^{t\pi}} \operatorname{Res}_{z=i\pi/2} \left(\frac{e^{-itz}}{\cosh z} \right),$$

where $\operatorname{Res}_{z=z_0} f(z)$ is the residue of f at a pole z_0.

A calculation shows that

$$\widehat{\varphi}(t) = \frac{\pi}{\cosh(t\pi/2)}.$$

5.6.5 Exercise

More generally consider the function

$$\varphi(x) = \frac{1}{\cosh x + a}, \quad -1 < a < 1. \tag{5.51}$$

Integrate the function

$$f(z) = \frac{e^{-itz}}{\cosh z + a}$$

along the rectangular contour with vertices $-R, R, R+i2\pi$ and $-R+i2\pi$. The function f has two simple poles $z = i(\pi \pm \arccos a)$ inside this rectangle. Proceed as in Exercise 5.6.4 to show

$$\widehat{\varphi}(t) = \frac{2\pi \sinh(t \arccos a)}{\sqrt{1-a^2} \sinh t\pi}. \tag{5.52}$$

It is plain that $\widehat{\varphi}(t) \geq 0$. Hence by Bochner's theorem $\varphi(x)$ is positive definite for $-1 < a < 1$. By a continuity argument it is positive definite for $a = 1$ as well.

5.6.6 Exercise

Now consider the function

$$\varphi(x) = \frac{1}{\cosh x + a}, \quad a > 1. \tag{5.53}$$

Use the function f and the rectangular contour of Exercise 5.6.5. Now f has two simple poles $z = \pm \operatorname{arccosh} t + i\pi$ inside this rectangle. Show that

$$\widehat{\varphi}(t) = \frac{2\pi \sin(t \operatorname{arccosh} a)}{\sqrt{a^2-1} \sinh t\pi}. \tag{5.54}$$

It is plain that $\widehat{\varphi}(t)$ is negative for some values of t. So the function φ in (5.53) is not positive definite for any $a > 1$.

5.6.7 Exercise

Let $\lambda_1, \ldots, \lambda_n$ be positive numbers and let Z be the $n \times n$ matrix with entries

$$z_{ij} = \frac{1}{\lambda_i^2 + \lambda_j^2 + t\lambda_i\lambda_j}.$$

Show that if $-2 < t \leq 2$, then Z is positive definite; and if $t > 2$ then there exists an $n > 2$ for which this matrix is not positive definite. (See Exercise 1.6.4.)

5.6.8 Exercise

For $0 < a < 1$, let f_a be the piecewise linear function defined as

$$f_a(x) = \begin{cases} 1 \ \text{ for } \ |x| \leq a, \\ 0 \ \text{ for } \ |x| \geq 1, \\ (1-a)^{-1}(1-|x|) \ \text{ for } \ a \leq |x| \leq 1. \end{cases}$$

Show that f_a is not positive definite. Compare this with 5.2.13 and 5.2.15. Express f_a as the convolution of two characteristic functions.

The technique introduced in Section 4 is a source of several interesting inequalities. The next two exercises illustrate this further.

5.6.9 Exercise

(i) Let A be a Hermitian matrix. Use the positive definiteness of the function sech x to show that for every matrix X

$$|||X||| \leq |||(I+A^2)^{1/2}X(I+A^2)^{1/2} - AXA|||.$$

(ii) Now let A be any matrix. Apply the result of (i) to the matrices $\tilde{A} = \begin{bmatrix} O & A \\ A^* & O \end{bmatrix}$ and $\widetilde{X} = \begin{bmatrix} O & X \\ X^* & O \end{bmatrix}$ and show that

$$|||X||| \leq |||(I+AA^*)^{1/2}X(I+A^*A)^{1/2} - AX^*A|||$$

for every matrix X. Replacing A by iA, one gets

$$|||X||| \leq |||(I+AA^*)^{1/2}X(I+A^*A)^{1/2} + AX^*A|||.$$

5.6.10 Exercise

Let A, B be normal matrices with $\|A\| \le 1$ and $\|B\| \le 1$. Show that for every X we have

$$|||(I - A^*A)^{1/2}X(I - B^*B)^{1/2}||| \le |||X - A^*XB|||.$$

The inequalities proved in Section 5.4 have a *leitmotiv.* Let $M(a, b)$ be any mean of positive numbers a and b satisfying the conditions laid down at the beginning of Chapter 4. Let A be a positive definite matrix with eigenvalues $\lambda_1 \ge \cdots \ge \lambda_n$. Let $M(A, A)$ be the matrix with entries

$$m_{ij} = M(\lambda_i, \lambda_j).$$

Many of the inequalites in Section 5.4 say that for certain means M_1 and M_2

$$\|M_1(A, A) \circ X\| \le \|M_2(A, A) \circ X\|, \tag{5.55}$$

for all X. We have proved such inequalites by showing that the matrix Y with entries

$$y_{ij} = \frac{M_1(\lambda_i, \lambda_j)}{M_2(\lambda_i, \lambda_j)} \tag{5.56}$$

is positive definite. This condition is also necessary for (5.55) to be true for all X.

5.6.11 Proposition

Let $M_1(a, b)$ and $M_2(a, b)$ be two means. Then the inequality (5.55) is true for all X if and only if the matrix Y defined by (5.56) is positive definite.

Proof. The Schur product by Y is a linear map on $\mathbb{M}_n$. The inequality (5.55) says that this linear map on the space $\mathbb{M}_n$ equipped with the norm $\|\cdot\|$ is contractive. Hence it is contractive also with respect to the dual norm $\|\cdot\|_1$; i.e.,

$$\|Y \circ X\|_1 \le \|X\|_1 \quad \text{for all } X.$$

Choose X to be the matrix with all entries equal to 1. This gives $\|Y\|_1 \le n$. Since Y is Hermitian

$$\|Y\|_1 = \sum_{i=1}^{n} |\lambda_i(Y)|$$

where $\lambda_i(Y)$ are the eigenvalues of Y. Since $y_{ii} = 1$ for all i, we have

$$\sum_{i=1}^{n} \lambda_i(Y) = \operatorname{tr} Y = n.$$

Thus $\sum |\lambda_i(Y)| \leq \sum \lambda_i(Y)$. But this is possible only if $\lambda_i(Y) \geq 0$ for all i. In other words Y is positive. ■

Let us say that $M_1 \leq M_2$ if $M_1(a,b) \leq M_2(a,b)$ for all positive numbers a and b; and $M_1 << M_2$ if for every n and every choice of n positive numbers $\lambda_1, \dots, \lambda_n$ the matrix (5.56) is positive definite. If $M_1 << M_2$ the inequality (5.55) is true for all positive matrices A and all matrices X. Clearly $M_1 \leq M_2$ if $M_1 << M_2$. The converse is not always true.

5.6.12 Exercise

Let $A(a,b)$ and $G(a,b)$ be the arithmetic and the geometric means of a and b. For $0 \leq \alpha \leq 1$ let

$$F_\alpha(a,b) = (1-\alpha)G(a,b) + \alpha A(a,b).$$

Clearly we have $F_\alpha \leq F_\beta$ whenever $\alpha \leq \beta$. Use Exercise 5.6.6 to show that $F_\alpha << F_\beta$ if and only if $\frac{1}{2} \leq \alpha \leq \beta$.

Using Exercise 2.7.12 one can see that if $M_1 << M_2$, then the inequality (5.55) is true for all unitarily invariant norms. The weaker condition $M_1 \leq M_2$ gives this inequality only for the Hilbert-Schmidt norm $\|\cdot\|_2$.

In Exercises 1.6.3, 1.6.4, 5.2.21, 5.2.22 and 5.2.23 we have outlined simple proofs of the infinite divisibility of some special matrices. These proofs rely on arguments specifically tailored to suit the matrices at hand. In the next few exercises we sketch proofs of some general theorems that are useful in this context.

An $n \times n$ Hermitian matrix A is said to be *conditionally positive definite* if $\langle x, Ax\rangle \geq 0$ for all $x \in \mathbb{C}^n$ such that $x_1 + \cdots + x_n = 0$. (The term *almost positive definite* is also used sometimes.)

5.6.13 Proposition

Let $A = [[a_{ij}]]$ be an $n \times n$ conditionally positive definite matrix. Then there exist a positive definite matrix $B = [[b_{ij}]]$ and a vector

$y = (y_1, \ldots, y_n)$ *in* $\mathbb{C}^n$ *such that*

$$a_{ij} = b_{ij} + y_i + \bar{y}_j. \tag{5.57}$$

Proof. Let J be the $n \times n$ matrix all of whose entries are equal to $1/n$. For any vector $x \in \mathbb{C}^n$ let $x^\# = Jx$ and $\tilde{x} = x - x^\#$. Since $\tilde{x}_1 + \cdots + \tilde{x}_n = 0$ we have

$$\langle \tilde{x}, A\tilde{x} \rangle \geq 0.$$

Let $B = A - AJ - JA + JAJ$. The inequality above says that

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

In other words B is positive definite.

If $C_1, \ldots, C_n$ are the columns of the matrix A, then the jth column of the matrix JA has all its entries equal to $\widehat{C}_j$, the number obtained by averaging the entries of the column C_j. Likewise, if $R_1, \ldots, R_n$ are the rows of A, then the ith row of AJ has all its entries equal to $\widehat{R}_i$. Since A is Hermitian $\widehat{R}_i$ is the complex conjugate of $\widehat{C}_i$. The matrix JAJ has all its entries equal to $\alpha = \frac{1}{n^2} \sum_{i,j} a_{ij}$. Thus the i, j entry of the matrix $AJ + JA - JAJ$ is equal to $\widehat{R}_i + \widehat{C}_j - \alpha$. Let y be the vector

$$y = \left(\widehat{R}_1 - \frac{\alpha}{2}, \widehat{R}_2 - \frac{\alpha}{2}, \ldots, \widehat{R}_n - \frac{\alpha}{2} \right).$$

Then the equation (5.57) is satisfied. ■

5.6.14 Exercise

Let $A = [[a_{ij}]]$ be a conditionally positive definite matrix. Show that the matrix $[[e^{a_{ij}}]]$ is positive definite. [Hint: If $B = [[b_{ij}]]$ is positive definite, then $[[e^{b_{ij}}]]$ is positive definite. Use Proposition 5.6.13.]

5.6.15 Exercise

Let $A = [[a_{ij}]]$ be an $n \times n$ matrix with positive entries and let $L = [[\log a_{ij}]]$. Let E be the matrix all whose entries are equal to 1. Note that $Ex = 0$ if $x_1 + \cdots + x_n = 0$.

(i) Suppose A is infinitely divisible. Let x be any vector with $x_1 + \cdots + x_n = 0$, and for $r > 0$ let $A^{(r)}$ be the matrix with entries a_{ij}^r. Then

$$\frac{1}{r}\langle x, (A^{(r)} - E)x \rangle = \frac{1}{r}\langle x, A^{(r)}x \rangle \geq 0.$$

Let $r \downarrow 0$. This gives

$$\langle x, Lx \rangle \geq 0.$$

Thus L is conditionally positive definite.

(ii) Conversely, if L is conditionally positive definite, then so is rL for every $r \geq 0$. Use Exercise 5.6.14 to show that this implies A is infinitely divisible.

Thus a Hermitian matrix A with entries $a_{ij} > 0$ is infinitely divisible if and only if the matrix $L = [[\log a_{ij}]]$ is conditionally positive definite. The next exercise gives a criterion for conditional positive definiteness.

5.6.16 Exercise

Given an $n \times n$ Hermitian matrix $B = [[b_{ij}]]$ let D be the $(n-1)\times(n-1)$ matrix with entries

$$d_{ij} = b_{ij} + b_{i+1,j+1} - b_{i,j+1} - b_{i+1,j}. \tag{5.58}$$

Show that B is conditionally positive definite if and only if D is positive definite.

We now show how the results of Exercises 5.6.14–5.6.16 may be used to prove the infinite divisibility of an interesting matrix.

For any n, the $n \times n$ *Pascal matrix* A is the matrix with entries

$$a_{ij} = \binom{i+j}{i} \quad \text{for } 0 \leq i, j \leq n-1. \tag{5.59}$$

The entries of the Pascal triangle occupy the antidiagonals of A. Thus the 4×4 Pascal matrix is

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \end{bmatrix}.$$

5.6.17 Exercise

Prove the combinatorial identity

$$\binom{i+j}{i} = \sum_{k=0}^{\min(i,j)} \binom{i}{k}\binom{j}{k}.$$

[Hint: Separate $i+j$ objects into two groups, the first containing i objects and the second j objects. If we choose $i-k$ objects from the first group and k from the second, we have chosen i objects out of $i+j$.]

5.6.18 Exercise

Show that

$$\binom{r+s}{r} = \frac{1}{2\pi}\int_0^{2\pi}(1+e^{i\theta})^r(1+e^{-i\theta})^s d\theta.$$

Use this to conclude that the Pascal matrix is a Gram matrix and is thus positive definite.

5.6.19 Exercise

Let L be the $n\times n$ lower triangular matrix whose rows are the rows of the Pascal triangle. Thus for $n=4$

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 1 & 3 & 3 & 1 \end{bmatrix}.$$

Show that $A = LL^*$. This gives another proof of the positive definiteness of the Pascal matrix A.

5.6.20 Exercise

For every n, the $n\times n$ Pascal matrix is infinitely divisible. Prove this statement following the steps given below.

(i) Use the results of Exercises 5.6.15 and 5.6.16. If B has entries $b_{ij} = \log a_{ij}$, where a_{ij} are defined by (5.59), then the entries d_{ij} defined by (5.58) are given by

$$d_{ij} = \log\left(1 + \frac{1}{i+j+1}\right).$$

We have to show that the matrix $D = [[d_{ij}]]$ is positive definite.

(ii) For $x > 0$ we have

$$\log(1+x) = \int_1^\infty \frac{tx}{t+x} d\mu(t)$$

where μ is the probability measure on $[0, \infty)$ defined as $d\mu(t) = dt/t^2$. Use this to show that

$$d_{ij} = \int_1^\infty \frac{1}{i+j+1+\frac{1}{t}} d\mu(t).$$

(iii) Thus the matrix D can be expressed as

$$D = \int_1^\infty C(t) d\mu(t),$$

where $C(t) = [[c_{ij}(t)]]$ is a Cauchy matrix for all $t \geq 1$. This shows that D is positive definite.

5.6.21 Exercise

The infinite divisibility of the Pascal matrix can be proved in another way as follows. Let $\lambda_1, \dots, \lambda_n$ be positive numbers, and let K be the $n \times n$ matrix with entries

$$k_{ij} = \frac{\Gamma(\lambda_i + \lambda_j + 1)}{\Gamma(\lambda_i + 1)(\lambda_j + 1)}.$$

When $\lambda_j = j, 1 \leq j \leq n$, this is the Pascal matrix. Use Gauss's product formula for the gamma function

$$\Gamma(z) = \lim_{m \to \infty} \frac{m! m^z}{z(z+1)\cdots(z+m)}, \quad z \neq 0, -1, -2, \dots,$$

to see that

$$k_{ij} = \lim_{m \to \infty} \frac{1}{m \cdot m!} \prod_{p=1}^{m+1} \frac{(\lambda_i + p)(\lambda_j + p)}{(\lambda_i + \lambda_j + p)}.$$

Each of the factors in this product is the i, j entry of a matrix that is congruent to a Cauchy matrix. Hence K is infinitely divisible.

Let f be a nonnegative function on $\mathbb{R}$. If for each $r > 0$ the function $(f(x))^r$ is positive definite, then we say that f is an *infinitely divisible function.* By Schur's theorem, the product of two infinitely divisible functions is infinitely divisible. If f is a nonnegative function and for each $m = 1, 2, \ldots$ the function $(f(x))^{1/m}$ is positive definite, then f is infinitely divisible.

Some examples of infinitely divisible functions are given in the next few exercises.

5.6.22 Exercise

(i) The function $f(x) = 1/(\cosh\, x)$ is infinitely divisible.

(ii) The function $f(x) = 1/(\cosh\, x + a)$ is infinitely divisible for $-1 < a \leq 1$. [Hint: Use Exercises 1.6.4 and 1.6.5.]

5.6.23 Exercise

In Section 5.2.10 we saw that the function

$$f(x) = \frac{\cosh\, \alpha\, x}{\cosh\, x}, \quad -1 \leq \alpha \leq 1,$$

is positive definite. Using this information and Schur's theorem one can prove that f is in fact infinitely divisible. The steps of the argument are outlined.

(i) Let a and b be any two nonnegative real numbers. Then

either $\dfrac{\cosh(a-b)x}{\cosh\, a\, x}$ or $\dfrac{\cosh(a-b)x}{\cosh\, b\, x}$ is positive definite.
Hence

$$\frac{\cosh(a-b)x}{\cosh\, ax\;\; \cosh\, bx}$$

is positive definite.

(ii) Use the identity

$$\cosh(a+b)x + \cosh(a-b)x = 2\cosh\, ax\;\; \cosh\, bx$$

to obtain

$$\frac{\cosh\, bx}{\cosh(a+b)x} = \frac{1}{2\cosh\, ax}\;\; \frac{1}{1 - \frac{\cosh(a-b)x}{2\cosh\, ax\;\; \cosh\, bx}}.$$

(iii) For $0 < r < 1$ we have the expansion

$$\left(\frac{\cosh\, bx}{\cosh(a+b)x}\right)^r = \frac{1}{2^r(\cosh\, ax)^r} \sum_{n=0}^{\infty} \frac{a_n}{2^n} \frac{\cosh^n(a-b)x}{\cosh^n\, ax\, \cosh^n\, bx},$$

where the coefficients a_n are nonnegative. Use Part (i) of this exercise and of Exercise 5.6.22 to prove that the series above represents a positive definite function. This establishes the assertion that $(\cosh \alpha x)/(\cosh\, x)$ is infinitely divisible for $0 < \alpha < 1$.

5.6.24 Exercise

The aim of this exercise is to show that the function

$$f(x) = \frac{\sinh \alpha x}{\sinh\, x}, \quad 0 < \alpha < 1,$$

is infinitely divisible. Its positive definiteness has been established in Section 5.2.10.

(i) Imitate the arguments in Exercise 5.6.23. Use the identity

$$\sinh(a+b)x + \sinh(a-b)x = 2\sinh\, ax\ \cosh\, bx$$

to show that the function

$$\frac{\sinh\, ax}{\sinh(a+b)x}$$

is infinitely divisible for $0 \le b \le a$. (This restriction is needed to handle the term $\sinh(a-b)x$ occurring in the series expansion.) This shows that the function $(\sinh\, \alpha x)/(\sinh\, x)$ is infinitely divisible for $1/2 \le \alpha \le 1$.

(ii) Let α be any number in $(0,1)$ and choose a sequence

$$\alpha = \alpha_0 < \alpha_1 < \alpha_2 < \cdots < \alpha_m = 1$$

where $\alpha_i/\alpha_{i+1} \ge 1/2$. Then

$$\frac{\sinh\, \alpha x}{\sinh\, x} = \prod_{i=0}^{m-1} \frac{\sinh\, \alpha_i x}{\sinh \alpha_{i+1} x}.$$

Each factor in this product is infinitely divisible, and hence so is the product.

5.6.25 Exercise

(i) Use Exercise 5.6.24 to show that the function $x/(\sinh x)$ is infinitely divisible. [Hint: Take limit $\alpha \downarrow 0$.]

(ii) Use this and the result of Exercise 5.6.23 to show that the function

$$f(x) = \frac{x \cosh \alpha x}{\sinh x}, \quad -1/2 \le \alpha \le 1/2,$$

is infinitely divisible.

5.6.26 Exercise

Let $\lambda_1, \ldots, \lambda_n$ be any real numbers. Use the result of Exercise 5.2.22 to show that the matrix

$$\left[\left[e^{-|\lambda_i - \lambda_j|}\right]\right]$$

is infinitely divisible. Thus the function $f(x) = e^{-|x|}$ is infinitely divisible. Use the integral formula

$$e^{-r|x|} = \frac{1}{r\pi} \int_{-\infty}^{\infty} \frac{e^{-itx}}{1 + t^2/r^2}\, dt, \quad r > 0,$$

to obtain another proof of this fact.

5.6.27 Exercise

Using the gamma function, as in Exercise 1.6.4, show that for every $r > 0$

$$\frac{1}{(1+ix)^r} = \frac{1}{\Gamma(r)} \int_0^{\infty} e^{-itx}\, e^{-t}\, t^{r-1}\, dt.$$

Thus the functions $1/(1+ix)^r$, $1/(1-ix)^r$, and $1/(1+x^2)^r$ are positive definite for every $r > 0$. This shows that $1/(1+x^2)$ is infinitely divisible.

5.6.28 Exercise

Let a and b be nonnegative numbers with $a \ge b$. Let $0 < r < 1$. Use the integral formula (1.39) to show that

$$\left(\frac{1+bx^2}{1+ax^2}\right)^r = \int_0^{\infty} \frac{1+bx^2}{1+\lambda+(a\lambda+b)x^2}\, d\mu(\lambda),$$

where μ is a positive measure. This is equal to

$$\int_0^\infty \left(\frac{b}{a\lambda + b} + \frac{\lambda(a-b)}{a\lambda + b} \; \frac{1}{1+\lambda+(a\lambda+b)x^2} \right) d\mu(\lambda).$$

Show that this is positive definite as a function of x. Note that it suffices to show that for each $\lambda > 0$,

$$g_\lambda(x) = \frac{1}{1+\lambda+(a\lambda+b)x^2}$$

is positive definite. This, in turn, follows from the integral representation

$$g_\lambda(x) = \frac{1}{2\gamma(1+\lambda)} \int_{-\infty}^{\infty} e^{-itx} e^{-|t|/\gamma} dt,$$

where $\gamma = [(a\lambda + b)/(1+\lambda)]^{1/2}$. Thus, for $a \geq b$ the function $f(x) = (1+bx^2)/(1+ax^2)$ is infinitely divisible.

5.6.29 Exercise

Show that the function $f(x) = (\tanh\, x)/x$ is infinitely divisible. [Hint: Use the infinite product expansion for $f(x)$.]

5.6.30 Exercise

Let $t > -1$ and consider the function

$$f(x) = \frac{\sinh\, x}{x(\cosh\, x + t)}.$$

Use the identity

$$\cosh\, x = 2\, \cosh^2 \frac{x}{2} - 1$$

to obtain the equality

$$f(x) = \frac{\sinh\,(x/2)\, \cosh(x/2)}{(x/2)\; 2\, \cosh^2(x/2)} \; \frac{1}{1 - \frac{1-t}{2\,\cosh^2(x/2)}}.$$

Use the binomial theorem and Exercise 5.6.29 to prove that f is infinitely divisible for $-1 < t \leq 1$.

Thus many of the positive definite functions from Section 5.2 are infinitely divisible. Consequently the associated positive definite matrices are infinitely divisible. In particular, for any positive numbers

$\lambda_1, \ldots, \lambda_n$ the $n \times n$ matrices V, W and Y whose entries are, respectively,

$$v_{ij} = \frac{\lambda_i^\alpha - \lambda_j^\alpha}{\lambda_i - \lambda_j}, \quad 0 < \alpha < 1,$$
$$w_{ij} = \frac{\log\, \lambda_i - \log\, \lambda_j}{\lambda_i - \lambda_j},$$
$$y_{ij} = \frac{\lambda_i^\nu + \lambda_j^\nu}{\lambda_i + \lambda_j}, \quad -1 \le \alpha \le 1,$$

are infinitely divisible.

5.6.31 Another proof of Bochner's Theorem

The reader who has worked her way through the theory of *Pick functions* (as given in Chapter V of MA) may enjoy the proof outlined below.

(i) Let φ be a positive definite function on $\mathbb{R}$, continuous at 0. Let $z = x + iy$ be a complex number and put

$$f(z) = \int_0^\infty e^{itz} \varphi(t) dt. \tag{5.60}$$

Since φ is bounded, this integral is convergent for $y > 0$. Thus f is an analytic function on the open upper half-plane H_+.

(ii) Observe that

$$\int_0^\infty e^{iv(z-\bar{z})} dv = \frac{1}{2y},$$

and so from (5.60) we have

$$\frac{\text{Re}\ f(z)}{y} = \int_0^\infty \int_0^\infty e^{i(t+v)z} e^{-iv\bar{z}} \varphi(t) dt\ dv + \int_0^\infty \int_0^\infty e^{-i(t+v)\bar{z}} e^{ivz} \varphi(-t) dt\ dv.$$

First substitute $u = t + v$ in both the integrals, and then interchange u and v in the second integral to obtain

$$\frac{\text{Re}\ f(z)}{y} = \int_0^\infty \left[\int_v^\infty e^{i(uz - v\bar{z})} \varphi(u - v) du \right] dv$$

$$+\int_0^\infty \left[\int_u^\infty e^{i(uz-v\bar{z})}\varphi(u-v)dv\right] du.$$

Observe that these two double integrals are over the quarter-planes $\{(u,v) : u \geq v \geq 0\}$ and $\{(u,v) : v \geq u \geq 0\}$, respectively. Hence

$$\begin{aligned}\frac{\text{Re } f(z)}{y} &= \int_0^\infty \int_0^\infty e^{i(uz-v\bar{z})}\varphi(u-v)du\ dv \\ &= \int_0^\infty \int_0^\infty \varphi(u-v)e^{i(u-v)x}e^{-(u+v)y}du\ dv.\end{aligned}$$

Since φ is a positive definite function, this integral is nonnegative. (Write it as a limit of Riemann sums each of which is nonnegative.)

Thus f maps the upper half-plane into the right half-plane. So $i\,f(z)$ is a Pick function.

(iii) For $\eta > 0$

$$|\eta f(i\eta)| \leq \int_0^\infty \eta e^{-t\eta} |\varphi(t)|\, dt \leq |\varphi(0)|\,.$$

Hence, by Problem V.5.9 of MA, there exists a finite positive measure μ on $\mathbb{R}$ such that

$$if(z) = \int_{-\infty}^\infty \frac{1}{\lambda - z} d\mu(\lambda).$$

(iv) Thus we have

$$\begin{aligned}f(z) &= \int_{-\infty}^\infty \frac{i}{-\lambda+z} d\mu(\lambda) \\ &= \int_{-\infty}^\infty \int_0^\infty e^{i(-\lambda+z)t}dt\ d\mu(\lambda) \\ &= \int_0^\infty \left[\int_{-\infty}^\infty e^{-i\lambda t}d\mu(\lambda)\right] e^{itz}dt.\end{aligned}$$

(v) Compare the expression for f in (5.60) with the one obtained in (iv) and conclude

$$\varphi(t) = \int_{-\infty}^\infty e^{-i\lambda t}d\mu(\lambda).$$

This is the assertion of Bochner's theorem.

5.7 NOTES AND REFERENCES

Positive definite functions have applications in almost every area of modern analysis. In 1907 C. Carathéodory studied functions with power series

$$f(z) = \frac{a_0}{2} + a_1 z + a_2 z^2 + \cdots,$$

and found necessary and sufficient conditions on the sequence $\{a_n\}$ in order that f maps the unit disk into the right half-plane. In 1911 O. Toeplitz observed that Carathéodory's condition is equivalent to (5.1). The connection with Fourier series and transforms has been pointed out in this chapter. In probability theory positive definite functions arise as characteristic functions of various distributions. See E. Lukacs, *Characteristic Functions,* Griffin, 1960, and R. Cuppens, *Decomposition of Multivariate Probabilities,* Academic Press, 1975. We mention just one more very important area of their application: the theory of group representations.

Let G be a locally compact topological group. A (continuous) complex-valued function φ on G is positive definite if for each $N = 1, 2, \ldots,$ the $N \times N$ matrix $\left[\left[\varphi\left(g_s^{-1} g_r\right)\right]\right]$ is positive for every choice of elements $g_0, \ldots, g_{N-1}$ from G. A unitary representation of G is a homomorphism $g \mapsto U_g$ from G into the group of unitary operators on a Hilbert space $\mathcal{H}$ such that for every fixed $x \in \mathcal{H}$ the map $g \mapsto U_g x$ from G into $\mathcal{H}$ is continuous. (This is called strong continuity.) It is easy to see that if U_g is a unitary representation of G in the Hilbert space $\mathcal{H}$, then for every $x \in \mathcal{H}$ the function

$$\varphi(g) = \langle x, U_g x \rangle \tag{5.61}$$

is positive definite on G. (This is a generalization of Exercise 5.6.1.) The converse is an important theorem of Gelfand and Raikov proved in 1943. This says that for every positive definite function φ on G there exist a Hilbert space $\mathcal{H}$, a unitary representation U_g of G in $\mathcal{H}$, and a vector $x \in \mathcal{H}$ such that the equation (5.61) is valid. This is one of the first theorems in the representation theory of infinite groups. One of its corollaries is that every locally compact group has sufficiently many irreducible unitary representations. More precisely, for every element g of G different from the identity, there exists an irreducible unitary representation of G for which U_g is not the identity operator.

An excellent survey of positive definite functions is given in J. Stewart, *Positive definite functions and generalizations, an historical sur-*

vey, Rocky Mountain J. Math., 6 (1976) 409–434. Among books, we recommend C. Berg, J.P.R. Christensen, and P. Ressel, *Harmonic Analysis on Semigroups,* Springer, 1984, and Z. Sasvári, *Positive Definite and Definitizable Functions* Akademie-Verlag, Berlin, 1994.

In Section 5.2 we have constructed a variety of examples using rather elementary arguments. These, in turn, are useful in proving that certain matrices are positive. The criterion in 5.2.15 is due to G. Pólya, *Remarks on characteristic functions,* Proc. Berkeley Symp. Math. Statist. & Probability, 1949, pp.115-123. This criterion is very useful as its conditions can be easily verified.

The ideas and results of Sections 5.2 and 5.3 are taken from the papers R. Bhatia and K. R. Parthasarathy, *Positive definite functions and operator inequalities,* Bull. London Math. Soc. 32 (2000) 214–228, H. Kosaki, *Arithmetic-geometric mean and related inequalities for operators,* J. Funct. Anal., 15 (1998) 429–451, F. Hiai and H. Kosaki, *Comparison of various means for operators,* ibid., 163 (1999) 300–323, and F. Hiai and H. Kosaki, *Means for matrices and comparison of their norms,* Indiana Univ. Math. J., 48 (1999) 899–936.

The proof of Theorem 5.3.1 given here is from R. Bhatia and K. B. Sinha, *Derivations, derivatives and chain rules,* Linear Algebra Appl., 302/303 (1999) 231–244. Theorem 5.3.3 was proved by K. Löwner (C. Loewner) in *Über monotone Matrixfunctionen,* Math. Z., 38 (1934) 177–216. Loewner then used this theorem to show that a function is operator monotone on the positive half-line if and only if it has an analytic continuation mapping the upper half-plane into itself. Such functions are characterized by certain integral representations, namely, f is operator monotone if and only if

$$f(t) = \alpha + \beta t + \int_0^\infty \frac{\lambda t}{\lambda + t}\, d\mu(t) \tag{5.62}$$

for some real numbers α and β with $\beta \geq 0$, and a positive measure μ that makes the integral above convergent. The connection between positivity of Loewner matrices and complex functions is made via Carathéodory's theorem (mentioned at the beginning of this section) and its successors. Following Loewner's work operator monotonicity of particular examples such as 5.3.5–5.3.8 was generally proved by invoking the latter two criteria (analytic continuation or integral representation). The more direct proofs based on the positivity of Loewner matrices given here are from the 2000 paper of Bhatia and Parthasarathy.

The inequality (5.24) and the more general (5.28) were proved in R. Bhatia and C. Davis, *More matrix forms of the arithmetic-*

geometric mean inequality, SIAM J. Matrix Anal. Appl., 14 (1993) 132–136. For the operator norm alone, the inequality (5.28) was proved by E. Heinz, *Beiträge zur Störungstheorie der Spektralzerlegung,* Math. Ann., 123 (1951) 415–438. The inequality (5.24) aroused considerable interest and several different proofs of it were given by various authors. Two of them, R. A. Horn, *Norm bounds for Hadamard products and the arithmetic-geometric mean inequality for unitarily invariant norms,* Linear Algebra Appl., 223/224 (1995) 355–361, and R. Mathias, *An arithmetic-geometric mean inequality involving Hadamard products,* ibid., 184 (1993) 71–78, observed that the inequality follows from the positivity of the matrix in (5.27). The papers by Bhatia-Parthasarathy and Kosaki cited above were motivated by extending this idea further. The two papers used rather similar arguments and obtained similar results. The program was carried much further in the two papers of Hiai and Kosaki cited above to obtain an impressive variety of results on means. The interested reader should consult these papers as well as the monograph F. Hiai and H. Kosaki, *Means of Hilbert Space Operators,* Lecture Notes in Mathematics Vol. 1820, Springer, 2003.

The theorems of Herglotz and Bochner concern the groups $\mathbb{Z}$ and $\mathbb{R}$. They were generalized to locally compact abelian groups by A. Weil, by D. A. Raikov, and by A. Powzner, in independent papers appearing almost together. Further generalizations (non-abelian or non-locally compact groups) exist. The original proof of Bochner's theorem appears in S. Bochner, *Vorlesungen über Fouriersche Integrale,* Akademie-Verlag, Berlin, 1932. Several other proofs have been published. The one given in Section 5.5 is taken from R. Goldberg, *Fourier Transforms,* Cambridge University Press, 1961, and that in Section 5.6 from N. I. Akhiezer and I. M. Glazman, *Theory of Linear Operators in Hilbert Space,* Dover, 1993 (reprint of original editions). A generalization to distributions is given in L. Schwartz, *Théorie des Distributions,* Hermann, 1954.

Integral representations such as the one given by Bochner's theorem are often viewed as a part of "Choquet Theory." Continuous positive definite functions $\varphi(x)$ such that $\varphi(0) = 1$ form a compact convex set; the family $\left\{e^{itx} : t \in \mathbb{R}\right\}$ is the set of extreme points of this convex set.

Exercise 5.6.1 is an adumbration of the connections between positive definite functions and spectral theory of operators. A basic theorem of M. H. Stone in the latter subject says that every unitary representation $t \mapsto U_t$ of $\mathbb{R}$ in a Hilbert space $\mathcal{H}$ is of the form $U_t = e^{itA}$ for some (possibly unbounded) self-adjoint operator A. (The operator A is bounded if and only if $\|U_t - I\| \to 0$ as $t \to 0$.) The theorems of Stone and Bochner can be derived from each other. See M. Reed and B. Simon, *Methods of Modern Mathematical Physics, Vols. I, II,* Academic Press, 1972, 1975, Chapters VIII, IX.

A sequence $\{a_n\}_{n=0}^{\infty}$ is of *positive type* if for every positive integer N, we have

$$\sum_{r,s=0}^{N-1} a_{r+s}\xi_r\bar{\xi}_s \geq 0 \tag{5.63}$$

for every finite sequence of complex numbers $\xi_0, \xi_1, \dots, \xi_{N-1}$. This is equivalent to the requirement that for each $N = 1, 2, \dots,$ the $N \times N$ matrix

$$\begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{N-1} \\ a_1 & a_2 & a_3 & \cdots & a_N \\ \vdots & \vdots & \vdots & & \vdots \\ a_{N-1} & a_N & a_{N+1} & \cdots & a_{2N-2} \end{bmatrix} \tag{5.64}$$

is positive. Compare these conditions with (5.1) and (5.2). (Matrices of the form (5.64) are called Hankel matrices while those of the form (5.2) are Toeplitz matrices.) A complex valued function φ on the positive half-line $[0, \infty)$ is of positive type if for each N the $N \times N$ matrix

$$[[\varphi\,(x_r + x_s)]] \tag{5.65}$$

is positive for every choice of $x_0, \dots, x_{N-1}$ in $[0, \infty)$. A theorem of Bernstein and Widder says that φ is of positive type if and only if there exists a positive measure μ on $[0, \infty)$ such that

$$\varphi(x) = \int_0^\infty e^{-tx} d\mu(t), \tag{5.66}$$

i.e., φ is the *Laplace transform* of a positive measure μ. Such functions are characterized also by being *completely monotone,* which, by definition, means that

$$(-1)^m \varphi^{(m)}(x) \geq 0, \quad m = 0, 1, 2, \dots.$$

See MA p.148 for the connection such functions have with operator monotone functions. The book of Berg, Christensen, and Ressel cited above is a good reference for the theory of these functions.

Our purpose behind this discussion is to raise a question. Suppose f is a function mapping $[0, \infty)$ into itself. Say that f is in the class

$\mathcal{L}_\pm$ if for each N the matrix

$$\left[\left[\frac{f(\lambda_i) \pm f(\lambda_j)}{\lambda_i \pm \lambda_j}\right]\right]$$

is positive for every choice $\lambda_1, \ldots, \lambda_N$ in $[0, \infty)$. The class $\mathcal{L}_-$ is precisely the operator monotone functions. Is there a good characterisation of functions in $\mathcal{L}_+$? One can easily see that if $f \in \mathcal{L}_+$, then so does $1/f$. It is known that $\mathcal{L}_-$ is contained in $\mathcal{L}_+$; see, e.g., M. K. Kwong, *Some results on matrix monotone functions,* Linear Algebra Appl., 118 (1989) 129–153. (It is easy to see, using the positivity of the Cauchy matrix, that for every $\lambda > 0$ the function $g(t) = \lambda t/(\lambda + t)$ is in $\mathcal{L}_+$. The integral representation (5.62) then shows that every function in $\mathcal{L}_-$ is in $\mathcal{L}_+$.)

The conjecture stated after Exercise 5.6.3 goes back to D. Bessis, P. Moussa, and M. Villani, *Monotonic converging variational approximations to the functional integrals in quantum statistical mechanics,* J. Math. Phys., 16 (1975) 2318–2325. A more recent report on the known partial results may be found in P. Moussa, *On the representation of Tr $\left(e^{(A-\lambda B)}\right)$ as a Laplace transform,* Rev. Math. Phys., 12 (2000) 621–655. E. H. Lieb and R. Seiringer, *Equivalent forms of the Bessis-Moussa-Villani conjecture,* J. Stat. Phys., 115 (2004) 185–190, point out that the statement of this conjecture is equivalent to the following: for all A and B positive, and all natural numbers p, the polynomial $\lambda \mapsto \operatorname{tr}(A + \lambda B)^p$ has only positive coefficients. When this polynomial is multiplied out, the co-efficient of λ^r is a sum of terms each of which is the trace of a word in A and B. It has been shown by C. R. Johnson and C. J. Hillar, *Eigenvalues of words in two positive definite letters,* SIAM J. Matrix Anal. Appl., 23 (2002) 916-928, that some of the individual terms in this sum can be negative. For example, $\operatorname{tr} A^2B^2AB$ can be negative even when A and B are positive.

The matrix Z in Exercise 5.6.7 was studied by M. K. Kwong, *On the definiteness of the solutions of certain matrix equations,* Linear Algebra Appl., 108 (1988) 177–197. It was shown here that for each $n \geq 2$, there exists a number t_n such that Z is positive for all t in $(-2, t_n]$, and further $t_n > 2$ for all n, $t_n = \infty, 8, 4$ for $n = 2, 3, 4$, respectively. The complete solution (given in Exercise 5.6.7) appears in the 2000 paper of Bhatia-Parthasarathy cited earlier. The idea and the method are carried further in R. Bhatia and D. Drissi, *Generalised Lyapunov equations and positive definite functions,* SIAM J. Matrix Anal. Appl., 27 (2005) 103-295–114. Using a Fourier transforms argument D. Drissi, *Sharp inequalities for some operator means,* preprint 2006, has shown that the function $f(x) = (x \cosh \alpha x)/\sinh x$ is not positive definite when $|\alpha| > 1/2$. The result of Exercise 5.6.9 is due to E. Andruchow, G. Corach, and D. Stojanoff, *Geometric operator inequalities,* Linear Algebra Appl., 258 (1997) 295-310, where other related inequalities are also discussed. The result of Exercise 5.6.10 was proved by D. K. Jocić, *Cauchy-Schwarz and means inequalities for elementary operators into norm ideals,* Proc. Am. Math. Soc., 126 (1998) 2705–2711. Cognate results are proved in D. K. Jocić, *Cauchy-Schwarz norm inequalities for weak*-integrals of operator valued functions,* J. Funct. Anal., 218 (2005) 318–346.

Proposition 5.6.11 is proved in the Hiai-Kosaki papers cited earlier. They also give an example of two means where $M_1 \leq M_2$, but $M_1 << M_2$ is not true. The simple example in Exercise 5.6.12 is from R. Bhatia, *Interpolating the arithmetic-geometric mean inequality and its operator version,* Linear Algebra Appl., 413 (2006) 355–363.

Conditionally positive definite matrices are discussed in Chapter 4 of the book R. B. Bapat and T. E. S. Raghavan, *Nonnegative Matrices and Applications,* Cambridge University Press, 1997, and more briefly in Section 6.3 of R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis,* Cambridge University Press, 1991. This section also contains a succinct discussion of infinitely divisible matrices and references to original papers. The results of Exercises 5.6.20 and 5.6.21 are taken from R. Bhatia, *Infinitely divisible matrices,* Am. Math. Monthly, 113 (2006) 221–235, and those of Exercises 5.6.23, 5.6.24 and 5.6.25 from R. Bhatia and H. Kosaki, *Mean matrices and infinite divisibility,* preprint 2006. In this paper it is shown that for several classes of means $m(a, b)$, matrices of the form $[[m(\lambda_i, \lambda_j)]]$ are infinitely divisible if $m(a, b) \leq \sqrt{ab}$ for all a and b; and if $\sqrt{ab} \leq m(a, b)$, then matrices of the form $[[1/m(\lambda_i, \lambda_j)]]$ are infinitely divisible. The contents of Exercise 5.6.28 are taken from H. Kosaki, *On infinite divisibility of positive definite functions,* preprint 2006. In this paper Kosaki uses very interesting ideas from complex analysis to obtain criteria for infinite divisibility.

We have spoken of Loewner's theorems that say that the Loewner matrices associated with a function f on $[0, \infty)$ are positive if and only if f has an analytic continuation mapping the upper half-plane into itself. R. A. Horn, *On boundary values of a schlicht mapping,* Proc. Am. Math. Soc., 18 (1967) 782–787, showed that this analytic continuation is a one-to-one (schlicht) mapping if and only if the Loewner matrices associated with f are infinitely divisible. This criterion gives another proof of the infinite divisibility of the matrices in Sections 5.3.5 and 5.3.7.

Infinitely divisible distribution functions play an important role in probability theory. These are exactly the limit distributions for sums of independent random variables. See, for example, L. Breiman, *Probability,* Addison-Wesley, 1968, pp. 190–196, or M. Loeve, *Probability Theory*, Van Nostrand, 1963, Section 22. The two foundational texts on this subject are B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables,* Addison-Wesley, 1954, and P. Lévy, *Théorie de l' Addition des Variables Aléatoires,* Gauthier-Villars, 1937. The famous Lévy-Khintchine Formula says that a continuous positive definite function φ, with $\varphi(0) = 1$, is infinitely divisible if and only if it can be represented as

$$\log \varphi(t) = i\gamma t + \int_{-\infty}^{\infty} \left(e^{itu} - 1 - \frac{itu}{1+u^2} \right) \frac{1+u^2}{u^2} dG(u),$$

where γ is a real number, G is a nondecreasing function of bounded variation, and the integrand at $u = 0$ is to be interpreted as the limiting value $-t^2/2$. This representation of $\log f(t)$ is unique. See the book by Gnedenko and Kolmogorov, p. 76. A list of such representations for some of the examples we have considered is given in Lukacs, p. 93.

A positive definite function is said to be *self-decomposable* if for every real number ν in $(0, 1)$ there exists a positive definite function φ_ν such that

$$\varphi(x) = \varphi(\nu x)\varphi_\nu(x).$$

Such functions are of importance in probability theory; see Section 23.3 of Loeve's book. It is shown there that if φ is self-decomposable, then it is infinitely divisible, and so are the functions φ_ν, $0 < \nu < 1$. Our discussion in Section 5.2.10 shows that the functions $1/(1 + x^2)$, $1/\cosh x$ and $x/\sinh x$ are self-decomposable. So, the infinite divisibility of some of the functions in Exercises 5.6.22–5.6.28 is a consequence also of Loeve's theorem.

In Exercise 5.6.28 we observed that if $a \geq b \geq 0$, then the function $f(x) = (1 + bx^2)/(1 + ax^2)$ is infinitely divisible, and in Exercise 5.6.29 asked the reader to use this to show that $g(x) = (\tanh x)/x$ is infinitely divisible. In the paper Z. J. Jurek and M. Yor, *Self decomposable laws associated with hyperbolic functions,* Probability and Mathematical Statistics, 24 (2004), 181–190, it is observed that the function f is not self-decomposable but the function g is. Some of the results in H. Kosaki, *On infinite divisibility of positive definite functions,* preprint 2006, can be rephrased to say that certain functions are self-decomposable. His list includes, for example, $1/(\cosh x + a)$, $-1 \leq a \leq 1$. The relevance and importance of these functions in probability theory is explained in the paper of Jurek and Yor, and in the references therein.

Finally, we make a few comments on operator inequalities with special reference to the arithmetic-geometric mean inequality. Operator inequalities are sought and found in three different versions. If A and B are positive, we may have sometimes an inequality $A \geq B$. By Weyl's monotonicity principle, this implies that $\lambda_j(A) \geq \lambda_j(B)$, $1 \leq j \leq n$, where $\lambda_j(A)$ is the jth eigenvalue of A counted in decreasing order. (This is equivalent to the existence of a unitary matrix U such that $A \geq UBU^*$.) This, in turn implies that $|||A||| \geq |||B|||$ for all unitarily invariant norms. If A and B are not positive, we may ask whether $|A| \geq |B|$. This implies the set of inequalities $s_j(A) \geq s_j(B)$ for all the singular values, which in turn implies $|||A||| \geq |||B|||$ for all unitarily invariant norms.

In Chapter 4 we saw an arithmetic-geometric mean inequality of the first kind; viz., $A\#B \leq \frac{1}{2}(A + B)$ for any two positive matrices. The inequality (5.23) and its stronger version (5.24) are of the third kind. An inequality of the second kind was proved by R. Bhatia and F. Kittaneh, *On the singular values of a product of operators,* SIAM J. Matrix Anal. Appl., 11 (1990) 272–277. This says that $s_j\left(A^{1/2}B^{1/2}\right) \leq s_j(A + B)/2$, for $1 \leq j \leq n$. This inequality for singular values implies the inequality (5.23) for norms. A stronger version $s_j(A^{1/2}XB^{1/2}) \leq s_j(AX + XB)/2$ is not always true. So, there is no second level inequality generalising (5.24).

For positive numbers a and b, the arithmetic-geometric mean inequality may be written in three different ways:

(i) $\sqrt{ab} \leq (a + b)/2$,

(ii) $ab \leq (a^2 + b^2)/2$,

(iii) $ab \leq ((a + b)/2)^2$.

While each of these three may be obtained from the other, the matrix versions suggested by them are different. For example (i) leads to the question whether

$$s_j^{1/2}(AB) \leq s_j\left(\frac{A+B}{2}\right)?$$

This is *different* from the inequality of Bhatia and Kittaneh stated above. It is not known whether this is true when $n > 2$. Weaker than this is the third-level inequality

$$||| \, |AB|^{1/2} \, ||| \leq \frac{1}{2} |||A + B|||.$$

This too is known to be true for a large class of unitarily invariant norms (including Schatten p-norms for $p = 1$ and for $p \geq 2$). It is not known whether it is true for all unitarily invariant norms. From properties of the matrix square function, one can see that this last (unproven) inequality is stronger than the assertion

$$|||AB||| \leq \left|\left|\left| \left(\frac{A+B}{2} \right)^2 \right|\right|\right|.$$

This version of the arithmetic-geometric mean inequality is known to be true. Thus there are quite a few subtleties involved in noncommutative versions of simple inequalities. A discussion of some of these matters may be found in R. Bhatia and F. Kittaneh, *Notes on matrix arithmetic-geometric mean inequalities,* Linear Algebra Appl., 308 (2000) 203–211, where the results just mentioned are proved. We recommend the monograph X. Zhan, *Matrix Inequalities*, Lecture Notes in Mathematics Vol. 1790, Springer, 2002 for a discussion of several topics related to these themes.

Chapter Six

Geometry of Positive Matrices

The set of $n \times n$ positive matrices is a differentiable manifold with a natural Riemannian structure. The geometry of this manifold is intimately connected with some matrix inequalities. In this chapter we explore this connection. Among other things, this leads to a deeper understanding of the geometric mean of positive matrices.

6.1 THE RIEMANNIAN METRIC

The space $\mathbb{M}_n$ is a Hilbert space with the inner product $\langle A, B\rangle = \operatorname{tr} A^*B$ and the associated norm $\|A\|_2 = (\operatorname{tr} A^*A)^{1/2}$. The set of Hermitian matrices constitutes a real vector space $\mathbb{H}_n$ in $\mathbb{M}_n$. The subset $\mathbb{P}_n$ consisting of *strictly* positive matrices is an open subset in $\mathbb{H}_n$. Hence it is a differentiable manifold. The tangent space to $\mathbb{P}_n$ at any of its points A is the space $T_A\mathbb{P}_n = \{A\} \times \mathbb{H}_n$, identified for simplicity, with $\mathbb{H}_n$. The inner product on $\mathbb{H}_n$ leads to a Riemannian metric on the manifold $\mathbb{P}_n$. At the point A this metric is given by the differential

$$ds = \|A^{-1/2}dAA^{-1/2}\|_2 = \left[\operatorname{tr}\left(A^{-1}dA\right)^2\right]^{1/2}. \tag{6.1}$$

This is a mnemonic for computing the length of a (piecewise) differentiable path in $\mathbb{P}_n$. If $\gamma : [a,b] \to \mathbb{P}_n$ is such a path, we define its length as

$$L(\gamma) = \int_a^b \|\gamma^{-1/2}(t)\gamma'(t)\gamma^{-1/2}(t)\|_2\, dt. \tag{6.2}$$

For each $X \in GL(n)$ the congruence transformation $\Gamma_X(A) = X^*AX$ is a bijection of $\mathbb{P}_n$ onto itself. The composition $\Gamma_X \circ \gamma$ is another differentiable path in $\mathbb{P}_n$.

6.1.1 Lemma

For each $X \in GL(n)$ and for each differentiable path γ

$$L\Big(\Gamma_X \circ \gamma\Big) = L(\gamma). \tag{6.3}$$

Proof. Using the definition of the norm $\|\cdot\|_2$ and the fact that $\operatorname{tr} XY = \operatorname{tr} YX$ for all X and Y we have for each t

$$\begin{aligned}
&\|\Big(X^*\gamma(t)X\Big)^{-1/2}\Big(X^*\gamma(t)X\Big)'\Big(X^*\gamma(t)X\Big)^{-1/2}\|_2 \\
&= \left[\operatorname{tr}\Big(X^*\gamma(t)X\Big)^{-1}\Big(X^*\gamma(t)X\Big)'\Big(X^*\gamma(t)X\Big)^{-1}\Big(X^*\gamma(t)X\Big)'\right]^{1/2} \\
&= \left[\operatorname{tr} X^{-1}\gamma^{-1}(t)\gamma'(t)\gamma^{-1}(t)\gamma'(t)X\right]^{1/2} \\
&= \left[\operatorname{tr} \gamma^{-1}(t)\gamma'(t)\gamma^{-1}(t)\gamma'(t)\right]^{1/2} \\
&= \|\gamma^{-1/2}(t)\gamma'(t)\gamma^{-1/2}(t)\|_2.
\end{aligned}$$

Intergrating over t we get (6.3). ■

For any two points A and B in $\mathbb{P}_n$ let

$$\delta_2(A,B) = \inf\ \{L(\gamma) : \gamma \text{ is a path from } A \text{ to } B\}. \tag{6.4}$$

This gives a metric on $\mathbb{P}_n$. The triangle inequality

$$\delta_2(A,B) \le \delta_2(A,C) + \delta_2(C,B)$$

is a consequence of the fact that a path γ_1 from A to C can be adjoined to a path γ_2 from C to B to obtain a path from A to B. The length of this latter path is $L(\gamma_1) + L(\gamma_2)$.

According to Lemma 6.1.1 each Γ_X is an isometry for the length L. Hence it is also an isometry for the metric δ_2; i.e.,

$$\delta_2\Big(\Gamma_X(A), \Gamma_X(B)\Big) = \delta_2(A,B), \tag{6.5}$$

for all A, B in $\mathbb{P}_n$ and X in $GL(n)$.

This observation helps us to prove several properties of δ_2. We will see that the infimum in (6.4) is attained at a unique path joining A and B. This path is called the *geodesic* from A to B. We will soon obtain an explicit formula for this geodesic and for its length. The following inequality called the *infinitesimal exponential metric increasing property* (IEMI) plays an important role. Following the notation introduced in Exercise 2.7.15 we write De^H for the derivative of the exponential map at a point H of $\mathbb{H}_n$. This is a linear map on $\mathbb{H}_n$ whose action is given as

$$De^H(K) = \lim_{t\to 0} \frac{e^{H+tK} - e^H}{t}.$$

6.1.2 Proposition (IEMI)

For all H and K in $\mathbb{H}_n$ we have

$$\|e^{-H/2} De^H(K) e^{-H/2}\|_2 \geq \|K\|_2. \tag{6.6}$$

Proof. Choose an orthonormal basis in which $H = \text{diag}\,(\lambda_1, \ldots, \lambda_n)$. By the formula (2.40)

$$De^H(K) = \left[\left[\frac{e^{\lambda_i} - e^{\lambda_j}}{\lambda_i - \lambda_j} k_{ij}\right]\right].$$

Therefore, the i, j entry of the matrix $e^{-H/2}\, D\, e^H(K)\, e^{-H/2}$ is

$$\frac{\sinh\,(\lambda_i - \lambda_j)/2}{(\lambda_i - \lambda_j)/2}\, k_{ij}.$$

Since $(\sinh\, x)/x \geq 1$ for all real x, the inequality (6.6) follows. ■

6.1.3 Corollary

Let $H(t)$, $a \leq t \leq b$ be any path in $\mathbb{H}_n$ and let $\gamma(t) = e^{H(t)}$. Then

$$L(\gamma) \geq \int_a^b \|H'(t)\|_2\, dt. \tag{6.7}$$

Proof. By the chain rule $\gamma'(t) = D\, e^{H(t)}\Big(H'(t)\Big)$. So the inequality (6.7) follows from the definition of $L(\gamma)$ given by (6.2) and the IEMI (6.6). ■

If $\gamma(t)$ is any path joining A and B in $\mathbb{P}_n$, then $H(t) = \log\ \gamma(t)$ is a path joining $\log A$ and $\log B$ in $\mathbb{H}_n$. The right-hand side of (6.7) is the length of this path in the Euclidean space $\mathbb{H}_n$. This is bounded below by the length of the straight line segment joining $\log A$ and $\log B$. Thus $L(\gamma) \geq \|\log A - \log B\|_2$, and we have the following important corollary called the *exponential metric increasing property* (EMI).

6.1.4 Theorem (EMI)

For each pair of points A, B in $\mathbb{P}_n$ we have

$$\delta_2(A, B) \geq \|\log A - \log B\|_2. \tag{6.8}$$

In other words for any two matrices H and K in $\mathbb{H}_n$

$$\delta_2(e^H, e^K) \geq \|H - K\|_2. \tag{6.9}$$

So the map

$$(\mathbb{H}_n, \|\cdot\|_2) \xrightarrow{\exp} (\mathbb{P}_n, \delta_2) \tag{6.10}$$

increases distances, or is metric increasing.

Our next proposition says that when A and B commute there is equality in (6.8). Further the exponential map carries the line segment joining $\log A$ and $\log B$ in $\mathbb{H}_n$ to the geodesic joining A and B in $\mathbb{P}_n$. A bit of notation will be helpful here. We write $[H, K]$ for the line segment

$$H(t) = (1-t)H + tK, \quad 0 \le t \le 1$$

joining two points H and K in $\mathbb{H}_n$. If A and B are two points in $\mathbb{P}_n$ we write $[A, B]$ for the geodesic from A to B. The existence of such a path is yet to be established. This is done first in the special case of commuting matrices.

6.1.5 Proposition

Let A and B be commuting matrices in $\mathbb{P}_n$. Then the exponential function maps the line segment $[\log A, \log B]$ in $\mathbb{H}_n$ to the geodesic $[A, B]$ in $\mathbb{P}_n$. In this case

$$\delta_2(A, B) = \|\log A - \log B\|_2.$$

Proof. We have to verify that the path

$$\gamma(t) = \exp\Big((1-t)\log A + t\log B\Big), \quad 0 \le t \le 1,$$

is the unique path of shortest length joining A and B in the space $(\mathbb{P}_n, \delta_2)$. Since A and B commute, $\gamma(t) = A^{1-t}B^t$ and $\gamma'(t) = (\log B - \log A)\,\gamma(t)$. The formula (6.2) gives in this case

$$L(\gamma) = \int_0^1 \|\log A - \log B\|_2 \, dt = \|\log A - \log B\|_2.$$

The EMI (6.7) says that no path can be shorter than this. So the path γ under consideration is one of shortest possible length.

Suppose $\widetilde{\gamma}$ is another path that joins A and B and has the same length as that of γ. Then $\widetilde{H}(t) = \log \widetilde{\gamma}(t)$ is a path that joins $\log A$ and $\log B$ in $\mathbb{H}_n$, and by Corollary 6.1.3 this path has length $\|\log A - \log B\|_2$. But in a Euclidean space the straight line segment is the unique shortest path between two points. So $\widetilde{H}(t)$ is a reparametrization of the line segment $[\log A, \log B]$. ■

Applying the reasoning of this proof to any subinterval $[0, a]$ of $[0, 1]$ we see that the parametrization

$$H(t) = (1-t) \log A + t \log B$$

of the line segment $[\log A, \log B]$ is the one that is mapped isometrically onto $[A, B]$ along the whole interval. In other words the natural parametrisation of the geodesic $[A, B]$ when A and B commute is given by

$$\gamma(t) = A^{1-t} B^t, \quad 0 \le t \le 1,$$

in the sense that $\delta_2\Big(A, \gamma(t)\Big) = t\delta_2(A, B)$ for each t. The general case is obtained from this with the help of the isometries Γ_X.

6.1.6 Theorem

Let A and B be any two elements of $\mathbb{P}_n$. Then there exists a unique geodesic $[A, B]$ joining A and B. This geodesic has a parametrization

$$\gamma(t) = A^{1/2}\Big(A^{-1/2} B A^{-1/2}\Big)^t A^{1/2}, \quad 0 \le t \le 1, \tag{6.11}$$

which is natural in the sense that

$$\delta_2(A, \gamma(t)) = t\, \delta_2(A, B) \tag{6.12}$$

for each t. Further, we have

$$\delta_2(A, B) = \| \log A^{-1/2} B A^{-1/2} \|_2. \tag{6.13}$$

Proof. The matrices I and $A^{-1/2} B A^{-1/2}$ commute. So the geodesic $\big[I, A^{-1/2} B A^{-1/2}\big]$ is naturally parametrized as

$$\gamma_0(t) = \Big(A^{-1/2} B A^{-1/2}\Big)^t.$$

Applying the isometry $\Gamma_{A^{1/2}}$ we obtain the path

$$\gamma(t) = \Gamma_{A^{1/2}}\Big(\gamma_0(t)\Big) = A^{1/2}\Big(A^{-1/2} B A^{-1/2}\Big)^t A^{1/2}$$

joining the points $\Gamma_{A^{1/2}}(I) = A$ and $\Gamma_{A^{1/2}}\Big(A^{-1/2} B A^{-1/2}\Big) = B$. Since $\Gamma_{A^{1/2}}$ is an isometry this path is the geodesic $[A, B]$. The equality

(6.12) follows from the similar property for $\gamma_0(t)$ noted earlier. Using Proposition 6.1.5 again we see that

$$\begin{aligned}\delta_2(A,B) &= \delta_2\Big(I, A^{-1/2}BA^{-1/2}\Big)\\ &= \|\log I - \log\Big(A^{-1/2}BA^{-1/2}\Big)\|_2\\ &= \|\log A^{-1/2}BA^{-1/2}\|_2. \quad \blacksquare\end{aligned}$$

Formula (6.13) gives an explicit representation for the metric δ_2 that we defined via (6.4). This is the *Riemannian metric* on the manifold $\mathbb{P}_n$. From the definition of the norm $\|\cdot\|_2$ we see that

$$\delta_2(A,B) = \Big(\sum_{i=1}^{n} \log^2 \lambda_i(A^{-1}B)\Big)^{1/2}, \tag{6.14}$$

where λ_i are the eigenvalues of the matrix $A^{-1}B$.

6.1.7 The geometric mean again

The expression (4.10) defining the geometric mean $A\#B$ now appears in a new light. It is the *midpoint* of the geodesic γ joining A and B in the space $(\mathbb{P}_n, \delta_2)$. This is evident from (6.11) and (6.12). The symmetry of $A\#B$ in the two arguments A and B that we deduced by indirect arguments in Section 4.1 is now revealed clearly: the midpoint of the geodesic $[A,B]$ is the same as the midpoint of $[B,A]$.

The next proposition supplements the information given by the EMI.

6.1.8 Proposition

If for some A, $B \in \mathbb{P}_n$, the identity matrix I lies on the geodesic $[A,B]$, then A and B commute, $[A,B]$ is the isometric image under the exponential map of a line segment through O in $\mathbb{H}_n$, and

$$\log B = -\frac{1-\xi}{\xi} \log A, \tag{6.15}$$

where $\xi = \delta_2(A,I)/\delta_2(A,B)$.

Proof. From Theorem 6.1.6 we know that

$$I = A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^{\xi} A^{1/2},$$

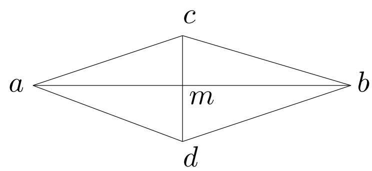
where $\xi = \delta_2\,(A,I)\,/\delta_2(A,B)$. Thus

$$B = A^{1/2}A^{-1/\xi}A^{1/2} = A^{-(1-\xi)/\xi}.$$

So A and B commute and (6.15) holds. Now Proposition 6.1.5 tells us that the exponential map sends the line segment $[\log A, \log B]$ isometrically onto the geodesic $[A, B]$. The line segment contains the point $O = \log I$. ■

While the EMI says that the exponential map (6.10) is metric nondecreasing in general, Proposition 6.1.8 says that this map is isometric on line segments through O. This essentially captures the fact that $\mathbb{P}_n$ is a Riemannian manifold of nonpositive curvature. See the discussion in Section 6.5.

Another essential feature of this geometry is the *semiparallelogram law* for the metric δ_2. To understand this recall the parallelogram law in a Hilbert space $\mathcal{H}$. Let a and b be any two points in $\mathcal{H}$ and let $m = (a + b)/2$ be their midpoint. Given any other point c consider the parallelogram one of whose diagonals is $[a, b]$ and the other $[c, d]$. The two diagonals intersect at m



and the parallelogram law is the equality

$$\|a - b\|^2 + \|c - d\|^2 = 2\Big(\|a - c\|^2 + \|b - c\|^2\Big).$$

Upon rearrangement this can be written as

$$\|c - m\|^2 = \frac{\|a - c\|^2 + \|b - c\|^2}{2} - \frac{\|a - b\|^2}{4}.$$

In the semiparallelogram law this last equality is replaced by an inequality.

6.1.9 Theorem (The Semiparallelogram Law)

Let A and B any two points of $\mathbb{P}_n$ and let $M = A\#B$ be the midpoint of the geodesic $[A, B]$. Then for any C in $\mathbb{P}_n$ we have

$$\delta_2^2(M, C) \leq \frac{\delta_2^2(A, C) + \delta_2^2(B, C)}{2} - \frac{\delta_2^2(A, B)}{4}. \tag{6.16}$$

Proof. Applying the isometry $\Gamma_{M^{-1/2}}$ to all matrices involved, we may assume that $M = I$. Now I is the midpoint of $[A, B]$ and so by Proposition 6.1.8 we have $\log B = -\log A$ and

$$\delta_2(A, B) = \|\log A - \log B\|_2.$$

The same proposition applied to $[M, C] = [I, C]$ shows that

$$\delta_2(M, C) = \|\log M - \log C\|_2.$$

The parallelogram law in the Hilbert space $\left(\mathbb{H}_n, \|.\|_2\right)$ tells us

$$\|\log M - \log C\|_2^2 = \frac{\|\log A - \log C\|_2^2 + \|\log B - \log C\|_2^2}{2} - \frac{\|\log A - \log B\|_2^2}{4}.$$

The left-hand side of this equation is equal to $\delta_2^2(M, C)$ and the subtracted term on the right-hand side is equal to $\delta_2^2(A, B)/4$. So the EMI (6.8) leads to the inequality (6.16). ■

In a Euclidean space the distance between the midpoints of two sides of a triangle is equal to half the length of the third side. In a space whose metric satisfies the semiparallelogram law this is replaced by an inequality.

6.1.10 Proposition

Let A, B, and C be any three points in $\mathbb{P}_n$. Then

$$\delta_2\left(A\#B, A\#C\right) \leq \frac{\delta_2(B, C)}{2}. \tag{6.17}$$

Proof. Consider the triangle with vertices A, B and C (and sides the geodesic segments joining the vertices). Let $M_1 = A\#B$. This is the midpoint of the side $[A, B]$ opposite the vertex C of the triangle $\{A, B, C\}$. Hence, by (6.16)

$$\delta_2^2\left(M_1, C\right) \leq \frac{\delta_2^2(A, C) + \delta_2^2(B, C)}{2} - \frac{\delta_2^2(A, B)}{4}.$$

Let $M_2 = A\#C$. In the triangle $\{A, M_1, C\}$ the point M_2 is the midpoint of the side $[A, C]$ opposite the vertex M_1. Again (6.16) tells us

$$\delta_2^2\left(M_1, M_2\right) \leq \frac{\delta_2^2(M_1, C) + \delta_2^2(M_1, A)}{2} - \frac{\delta_2^2(A, C)}{4}.$$

Substituting the first inequality into the second we obtain

$$\delta_2^2(M_1, M_2) \le \frac{1}{4}\left[\delta_2^2(A,C) + \delta_2^2(B,C)\right] - \frac{1}{8}\delta_2^2(A,B) + \frac{1}{2}\delta_2^2(M_1, A) - \frac{1}{4}\delta_2^2(A,C).$$

Since $\delta_2(M_1, A) = \delta_2(A,B)/2$, the right-hand side of this inequality reduces to $\delta_2^2(B,C)/4$. This proves (6.17). ■

The inequality (6.17) can be used to prove a more general version of itself. For $0 \le t \le 1$ let

$$A\#_t B = A^{1/2}\left(A^{-1/2} B A^{-1/2}\right)^t A^{1/2}. \tag{6.18}$$

This is another notation for the geodesic curve $\gamma(t)$ in (6.11). When $t = 1/2$ this is the geometric mean $A\#B$. The more general version is in the following.

6.1.11 Corollary

Given four points B, C, B', and C' in $\mathbb{P}_n$ let

$$f(t) = \delta_2\left(B'\#_t B,\ C'\#_t C\right).$$

Then f is convex on $[0,1]$; i.e.,

$$\delta_2\left(B'\#_t B,\ C'\#_t C\right) \le (1-t)\delta_2\left(B', C'\right) + t\delta_2(B,C). \tag{6.19}$$

Proof. Since f is continuous it is sufficient to prove that it is midpoint-convex. Let $M_1 = B'\#B$, $M_2 = C'\#C$, and $M = B'\#C$. By Proposition 6.1.10 we have $\delta_2(M_1, M) \le \delta_2(B,C)/2$ and $\delta_2(M, M_2) \le \delta_2(B', C')/2$. Hence

$$\delta_2(M_1, M_2) \le \delta_2(M_1, M) + \delta_2(M, M_2) \le \frac{1}{2}\left[\delta_2(B,C) + \delta_2(B', C')\right].$$

This shows that f is midpoint-convex. ■

Choosing $B' = C' = A$ in (6.19) gives the following theorem called the *convexity of the metric* δ_2.

6.1.12 Theorem

Let A, B and C be any three points in $\mathbb{P}_n$. Then for all t in $[0,1]$ we have

$$\delta_2\Big(A\#_t B, A\#_t C\Big) \le t\delta_2(B, C). \tag{6.20}$$

6.1.13 Exercise

For a fixed A in $\mathbb{P}_n$ let f be the function $f(X) = \delta_2^2(A, X)$. Show that if $X_1 \neq X_2$, then for $0 < t < 1$

$$f\Big(X_1\#_t X_2\Big) < (1-t)f(X_1) + tf(X_2). \tag{6.21}$$

This is expressed by saying that the function f is strictly convex on $\mathbb{P}_n$. [Hint: Show this for $t = 1/2$ first.]

6.2 THE METRIC SPACE $\mathbb{P}_n$

In this section we briefly study some properties of the metric space $(\mathbb{P}_n, \delta_2)$ with special emphasis on convex sets.

6.2.1 Lemma

The exponential is a continuous map from the space $(\mathbb{H}_n, \|.\|_2)$ onto the space $(\mathbb{P}_n, \delta_2)$.

Proof. Let H_m be a sequence in $\mathbb{H}_n$ converging to H. Then $e^{-H_m}e^H$ converges to I in the metric induced by $\|.\|_2$. So all the eigenvalues $\lambda_i\Big(e^{-H_m}e^H\Big)$, $1 \le i \le n$, converge to 1. The relation (6.14) then shows that $\delta_2\Big(e^{H_m}, e^H\Big)$ goes to zero as m goes to ∞. ■

6.2.2 Proposition

The metric space $(\mathbb{P}_n, \delta_2)$ is complete.

Proof. Let $\{A_m\}$ be a Cauchy sequence in $(\mathbb{P}_n, \delta_2)$ and let $H_m = \log A_m$. By the EMI (6.8) $\{H_m\}$ is a Cauchy sequence in $(\mathbb{H}_n, \|\cdot\|_2)$, and hence it converges to some H in $\mathbb{H}_n$. By Lemma 6.2.1 the sequence $\{A_m\}$ converges to $A = e^H$ in the space $(\mathbb{P}_n, \delta_2)$. ■

Note that $\mathbb{P}_n$ is *not* a complete subspace of $(\mathbb{H}_n, \|.\|_2)$. There it has a boundary consisting of singular positive matrices. In terms of the metric δ_2 these are "points at infinity." The next proposition shows that we may approach these points along geodesics. We use $A\#_t B$ for the matrix defined by (6.18) for every real t. When A and B commute, this reduces to $A^{1-t}B^t$.

6.2.3 Proposition

Let S be a singular positive matrix. Then there exist commuting elements A and B in $\mathbb{P}_n$ such that

$$\|A^{1-t}B^t - S\|_2 \to 0 \quad \text{and} \quad \delta_2\Big(A^{1-t}B^t, A\Big) \to \infty$$

as $t \to \infty$.

Proof. Apply a unitary conjugation and assume $S = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ where λ_k are nonnegative for $1 \le k \le n$, and $\lambda_k = 0$ for some k. If $\lambda_k > 0$, then put $\alpha_k = \beta_k = \lambda_k$, and if $\lambda_k = 0$, then put $\alpha_k = 1$ and $\beta_k = 1/2$. Let $A = \operatorname{diag}(\alpha_1, \dots, \alpha_n)$ and $B = \operatorname{diag}(\beta_1, \dots, \beta_n)$. Then

$$\lim_{t \to \infty} \|A^{1-t}B^t - S\|_2 = 0.$$

For the metric δ_2 we have

$$\begin{aligned}\delta_2\Big(A^{1-t}B^t, A\Big) &= \|\log A^{-1}A^{1-t}B^t\|_2 = \|\log A^{-t}B^t\|_2 \\ &\ge \left|\log 2^{-t}\right| = t \log 2,\end{aligned}$$

and this goes to ∞ as $t \to \infty$. ■

The point of the proposition is that the curve $A\#_t B$ starts at A when $t = 0$, and "goes away to infinity" in the metric space $(\mathbb{P}_n, \delta_2)$ while converging to S in the space $(\mathbb{H}_n, \|\cdot\|_2)$.

It is conventional to extend some matrix operations from strictly positive matrices to singular positive matrices by taking limits. For example, the geometric mean $A\#B$ is defined by (4.10) for strictly positive matrices A and B, and then defined for singular positive matrices A and B as

$$A\#B = \lim_{\varepsilon \downarrow 0} \Big(A + \varepsilon I\Big)\#\Big(B + \varepsilon I\Big).$$

The next exercise points to the need for some caution when using this idea.

6.2.4 Exercise

The geometric mean $A\#B$ is continuous on pairs of strictly positive matrices, but is not so when extended to positive semidefinite matrices. (See Exercise 4.1.6.)

We have seen that any two points A and B in $\mathbb{P}_n$ can be joined by a geodesic segment $[A, B]$ lying in $\mathbb{P}_n$. We say a subset $\mathcal{K}$ of $\mathbb{P}_n$ is *convex* if for each pair of points A and B in $\mathcal{K}$ the segment $[A, B]$ lies entirely in $\mathcal{K}$. If $\mathcal{S}$ is any subset of $\mathbb{P}_n$, then the *convex hull* of $\mathcal{S}$ is the smallest convex set containing $\mathcal{S}$. This set, denoted as $\operatorname{conv}(\mathcal{S})$ is the intersection of all convex sets that contain $\mathcal{S}$. Clearly, the convex hull of any two point set $\{A, B\}$ is $[A, B]$.

6.2.5 Exercise

Let $\mathcal{S}$ be any set in $\mathbb{P}_n$. Define inductively the sets $\mathcal{S}_m$ as $\mathcal{S}_0 = \mathcal{S}$ and

$$\mathcal{S}_{m+1} = \cup \{[A, B] : A, B \in \mathcal{S}_m\}.$$

Show that

$$\operatorname{conv}(\mathcal{S}) = \bigcup_{m=0}^{\infty} \mathcal{S}_m.$$

The next theorem says that if $\mathcal{K}$ is a closed convex set in $(\mathbb{P}_n, \delta_2)$, then a *metric projection* onto $\mathcal{K}$ exists just as it does in a Hilbert space.

6.2.6 Theorem

Let $\mathcal{K}$ be a closed convex set in $(\mathbb{P}_n, \delta_2)$. Then for each $A \in \mathbb{P}_n$ there exists a point $C \in \mathcal{K}$ such that $\delta_2(A, C) < \delta_2(A, K)$ for every K in $\mathcal{K}$, $K \neq C$. (In other words C is the unique best approximant to A from the set $\mathcal{K}$.)

Proof. Let $\mu = \inf\{\delta_2(A, K) : K \in \mathcal{K}\}$. Then there exists a sequence $\{C_n\}$ in $\mathcal{K}$ such that $\delta_2(A, C_n) \to \mu$. Given n and m, let M be the midpoint of the geodesic segment $[C_n, C_m]$; i.e., $M = C_n \# C_m$. By the convexity of $\mathcal{K}$ the point M is in $\mathcal{K}$. Using the semiparallelogram law (6.16) we get

$$\delta_2^2(M, A) \leq \frac{\delta_2^2(C_n, A) + \delta_2^2(C_m, A)}{2} - \frac{\delta_2^2(C_n, C_m)}{4},$$

and hence

$$\delta_2^2(C_n, C_m) \le 2\left[\delta_2^2(C_n, A) + \delta_2^2(C_m, A)\right] - 4\mu^2. \tag{6.22}$$

As n and m go to ∞, the right-hand side of (6.22) goes to zero. Hence $\{C_n\}$ is a Cauchy sequence, and by Proposition 6.2.2 it converges to a limit C in $\mathbb{P}_n$. Since $\mathcal{K}$ is closed, C is in $\mathcal{K}$. Further $\delta_2(A, C) = \lim \delta_2(A, C_n) = \mu$. If K is any other element of $\mathcal{K}$ such that $\delta_2(A, K) = \mu$, then putting $C_n = C$ and $C_m = K$ in (6.22) we see that $\delta_2(C, K) = 0$; i.e., $C = K$. ■

The map $\pi(A) = C$ given by Proposition 6.2.6 may be called the *metric projection* onto $\mathcal{K}$.

6.2.7 Theorem

Let π be the metric projection onto a closed convex set $\mathcal{K}$ of $\mathbb{P}_n$. If A is any point of $\mathbb{P}_n$ and $\pi(A) = C$, then for any D in $\mathcal{K}$

$$\delta_2^2(A, D) \ge \delta_2^2(A, C) + \delta_2^2(C, D). \tag{6.23}$$

Proof. Let $\{M_n\}$ be the sequence defined inductively as $M_0 = D$, and $M_{n+1} = M_n \# C$. Then $\delta_2(C, M_n) = 2^{-n}\delta_2(C, D)$, and M_n converges to $C = M_\infty$. By the semiparallelogram law (6.16)

$$2\delta_2^2(A, M_{n+1}) \le \delta_2^2(A, M_n) + \delta_2^2(A, C) - \frac{1}{2}\delta^2(C, M_n).$$

Hence,

$$\delta_2^2(A, M_n) - \delta_2^2(A, M_{n+1}) \ge \frac{1}{2\cdot 4^n}\delta_2^2(C, D) + \delta_2^2(A, M_{n+1}) - \delta_2^2(A, C).$$

Summing these inequalities we have

$$\sum_{n=0}^{\infty}\left[\delta_2^2(A, M_n) - \delta_2^2(A, M_{n+1})\right]$$
$$\ge \frac{2}{3}\delta_2^2(C, D) + \sum_{n=0}^{\infty}\left[\delta_2^2(A, M_{n+1}) - \delta_2^2(A, C)\right].$$

It is easy to see that the two series are absolutely convergent.

Let $d_n = \delta_2^2(A, M_n) - \delta_2^2(A, C)$. Then the last inequality can be written as

$$\delta_2^2(A, D) - \delta_2^2(A, C) = d_0 \ \ge\ \frac{2}{3}\delta_2^2(C, D) + \sum_{n=1}^{\infty} d_n.$$

The same argument applied to M_n in place of D shows

$$\delta_2^2(A, M_n) - \delta_2^2(A, C) = d_n \;\geq\; \frac{2}{3}\delta_2^2(C, M_n) + \sum_{k=n+1}^{\infty} d_k.$$

Thus

$$\begin{aligned}
d_0 &\geq \frac{2}{3}\delta_2^2(C, D) + d_1 + \sum_{k=2}^{\infty} d_k \\
&\geq \frac{2}{3}\delta_2^2(C, D) + \frac{2}{3}\delta_2^2(C, M_1) + 2\sum_{k=2}^{\infty} d_k \\
&= \frac{2}{3}(1 + \frac{1}{4})\delta_2^2(C, D) + 2d_2 + 2\sum_{k=3}^{\infty} d_k \\
&\geq \frac{2}{3}(1 + \frac{1}{4})\delta_2^2(C, D) + 2\left[\frac{2}{3}\delta_2^2(C, M_2) + \sum_{k=3}^{\infty} d_k\right] + 2\sum_{k=3}^{\infty} d_k \\
&= \frac{2}{3}\left(1 + \frac{1}{4} + \frac{2}{4^2}\right)\delta_2^2(C, D) + 4\sum_{k=3}^{\infty} d_k \\
&\geq \cdots.
\end{aligned}$$

Since $\mathcal{K}$ is convex, each $M_n \in \mathcal{K}$, and hence $d_n \geq 0$. Thus we have

$$d_0 \geq \frac{2}{3}\left[1 + \sum_{n=1}^{\infty} \frac{2^{n-1}}{4^n}\right]\delta_2^2(C, D) = \delta_2^2(C, D).$$

This proves the inequality (6.23). ■

6.2.8 The geometric mean once again

If $\mathcal{E}$ is a Euclidean space with metric d, and a, b are any two points of $\mathcal{E}$, then the function

$$f(x) = d^2(a, x) + d^2(b, x)$$

attains its minimum on $\mathcal{E}$ at the unique point $x_0 = \frac{1}{2}(a + b)$. In the metric space $(\mathbb{P}_n, \delta_2)$ this role is played by the geometric mean.

Proposition. *Let A and B be any two points of $\mathbb{P}_n$, and let*

$$f(X) = \delta_2^2(A, X) + \delta_2^2(B, X).$$

Then the function f is strictly convex on $\mathbb{P}_n$, and has a unique minimum at the point $X_0 = A\#B$.

Proof. The strict convexity is a consequence of Exercise 6.1.13. The semiparallelogram law implies that for every X we have

$$\delta_2^2\,(A\#B, X) \le \frac{1}{2}f(X) - \frac{1}{4}\delta_2^2(A,B) = \frac{1}{2}f(X) - \frac{1}{2}f(A\#B).$$

Hence

$$f(A\#B) \le f(X) - 2\delta_2^2(A\#B, X).$$

This shows that f has a unique minimum at the point $X_0 = A\#B$. ■

6.3 CENTER OF MASS AND GEOMETRIC MEAN

In Chapter 4 we discussed, and resolved, the problems associated with defining a good geometric mean of two positive matrices. In this section we consider the question of a suitable definition of a geometric mean of more than two matrices. Our discussion will show that while the case of two matrices is very special, ideas that work for three matrices do work for more than three as well.

Given three positive matrices A_1, A_2, and A_3, their geometric mean $G(A_1, A_2, A_3)$ should be a positive matrix with the following properties. If A_1, A_2, and A_3 commute with each other, then $G(A_1A_2A_3) = (A_1A_2A_3)^{1/3}$. As a function of its three variables, G should satisfy the conditions:

(i) $G(A_1, A_2, A_3) = G(A_{\pi(1)}, A_{\pi(2)}, A_{\pi(3)})$ for every permutation π of $\{1, 2, 3\}$.

(ii) $G(A_1, A_2, A_3) \le G(A_1', A_2, A_3)$ whenever $A_1 \le A_1'$.

(iii) $G(X^*A_1X, X^*A_2X, X^*A_3X) = X^*G(A_1, A_2, A_3)X$ for all $X \in GL(n)$.

(iv) G is continuous.

The first three conditions may be called *symmetry, monotonicity,* and *congruence invariance,* respectively.

None of the procedures that we used in Chapter 4 to define the geometric mean of two positive matrices extends readily to three. While two positive matrices can be diagonalized simultaneously by a congruence, in general three cannot be. The formula (4.10) has no obvious analogue for three matrices; nor does the extremal characterization (4.15). It is here that the connections with geometry made in Sections 6.1.7 and 6.2.8 suggest a way out: the geometric mean of three matrices should be the "center" of the triangle that has the three matrices as its vertices.

As motivation, consider the arithmetic mean of three points x_1, x_2, and x_3 in a Euclidean space $(\mathcal{E}, d)$. The point $\bar{x} = \frac{1}{3}(x_1 + x_2 + x_3)$ is characterized by several properties; three of them follow:

(i) $\bar{x}$ is the unique point of intersection of the three medians of the triangle $\triangle(x_1, x_2, x_3)$. (This point is called the *centroid* of $\triangle$.)

(ii) $\bar{x}$ is the unique point in $\mathcal{E}$ at which the function

$$d^2(x, x_1) + d^2(x, x_2) + d^2(x, x_3)$$

attains its minimum. (This point is the *center of mass* of the triple $\{x_1, x_2, x_3\}$ if each of them has equal mass.)

(iii) $\bar{x}$ is the unique point of intersection of the nested sequence of triangles $\{\triangle_n\}$ in which $\triangle_1 = \triangle(x_1, x_2, x_3)$ and $\triangle_{j+1}$ is the triangle whose vertices are the midpoints of the three sides of $\triangle_j$.

We may try to mimic these constructions in the space $(\mathbb{P}_n, \delta_2)$. As we will see, this has to be done with some circumspection.

The first difficulty is with the identification of a triangle in this space. In Section 6.2 we defined convex hulls and observed that the convex hull of two points A_1, A_2 in $\mathbb{P}_n$ is the geodesic segment $[A_1, A_2]$. It is harder to describe the convex hull of *three* points A_1, A_2, A_3. (This seems to be a difficult problem in Riemannian geometry.) In the notation of Exercise 6.2.5, if $\mathcal{S} = \{A_1, A_2, A_3\}$, then $\mathcal{S}_1 = [A_1, A_2] \cup [A_2, A_3] \cup [A_3, A_1]$ is the union of the three "edges." However, $\mathcal{S}_2$ is not in general a "surface," but a "fatter" object. Thus it may happen that the three "medians" $[A_1, A_2 \# A_3], [A_2, A_1 \# A_3]$, and $[A_3, A_1 \# A_2]$ do not intersect at all in most cases. So, we have to abandon this as a possible definition of the centroid of the triangle $\triangle(A_1, A_2, A_3)$.

Next we ask whether for every triple of points A_1, A_2, A_3 in $\mathbb{P}_n$ there exists a (unique) point X_0 at which the function

$$f(X) = \sum_{j=1}^{3} \delta_2^2(A_j, X)$$

attains its minimum value on $\mathbb{P}_n$. A simple argument using the semi-parallelogram law shows that such a point exists. This goes as follows.

Let $m = \inf f(X)$ and let $\{X_r\}$ be a sequence in $\mathbb{P}_n$ such that $f(X_r) \to m$. By the semiparallellgram law we have for $j = 1, 2, 3,$ and for all r and s

$$\delta_2^2\left(X_r \# X_s, A_j\right) \leq \frac{\delta_2^2(X_r, A_j) + \delta_2^2(X_s, A_j)}{2} - \frac{\delta_2^2(X_r, X_s)}{4}.$$

Summing up these three inequalities over j, we obtain

$$f\left(X_r \# X_s\right) \leq \frac{1}{2}\Big(f(X_r) + f(X_s)\Big) - \frac{3}{4}\delta_2^2(X_r, X_s).$$

This shows that

$$\begin{aligned}\frac{3}{4}\delta_2^2(X_r, X_s) &\leq \frac{1}{2}\Big(f(X_r) + f(X_s)\Big) - f\left(X_r \# X_s\right) \\ &\leq \frac{1}{2}\Big(f(X_r) + f(X_s)\Big) - m.\end{aligned}$$

It follows that $\{X_r\}$ is a Cauchy sequence, and hence it converges to a limit X_0. Clearly f attains its minimum at X_0. By Exercise 6.1.13 the function f is strictly convex and its minimum is attained at a unique point.

We define the "center of mass" of $\{A_1, A_2, A_3\}$ as the point

$$G(A_1, A_2, A_3) = \operatorname{arcmin} \sum_{j=1}^{3} \delta_2^2(A_j, X), \tag{6.24}$$

where the notation $\operatorname{arcmin} f(X)$ stands for the point X_0 at which the function $f(X)$ attains its minimum value. It is clear from the definition that $G(A_1, A_2, A_3)$ is a symmetric and continuous function of the three variables. Since each congruence transformation Γ_X is an isometry of $(\mathbb{P}_n, \delta_2)$ it is easy to see that G is congruence invariant; i.e.,

$$G(X^*A_1X, X^*A_2X, X^*A_3X) = X^*G(A_1, A_2, A_3)X.$$

Thus G has three of the four desirable properties listed for a good geometric mean at the beginning of this section. We do not know whether G is monotone. Some more properties of G are derived below.

6.3.1 Lemma

Let φ_1, φ_2 be continuously differentiable real-valued functions on the interval $(0, \infty)$ and let

$$h(X) = \Big\langle \varphi_1(X), \varphi_2(X) \Big\rangle = \operatorname{tr} \varphi_1(X)\varphi_2(X),$$

for all $X \in \mathbb{P}_n$. Then the derivative of h is given by the formula

$$Dh(X)(Y) = \Big\langle \varphi_1'(X)\varphi_2(X) + \varphi_1(X)\varphi_2'(X), Y \Big\rangle.$$

Proof. By the product rule for differentiation (see MA, p. 312) we have

$$Dh(X)(Y) = \Big\langle D\varphi_1(X)(Y), \varphi_2(X) \Big\rangle + \Big\langle \varphi_1(X), D\varphi_2(X)(Y) \Big\rangle.$$

Choose an orthonormal basis in which $X = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Then by (2.40)

$$D\varphi_1(X)(Y) = \left[\left[\frac{\varphi_1(\lambda_i) - \varphi_1(\lambda_j)}{\lambda_i - \lambda_j}\right]\right] \circ Y.$$

Hence,

$$\begin{aligned}\Big\langle D\varphi_1(X)(Y), \varphi_2(X) \Big\rangle &= \sum_i \varphi_1'(\lambda_i) y_{ii} \varphi_2(\lambda_i) \\ &= \Big\langle \varphi_1'(X)\varphi_2(X), Y \Big\rangle.\end{aligned}$$

Similarly,

$$\Big\langle \varphi_1(X), D\varphi_2(X)(Y) \Big\rangle = \Big\langle \varphi_1(X)\varphi_2'(X), Y \Big\rangle.$$

This proves the lemma. ■

6.3.2 Corollary

Let $h(X) = \|\log X\|_2^2, \quad X \in \mathbb{P}_n$. Then

$$Dh(X)(Y) = 2\Big\langle X^{-1}\log X, Y \Big\rangle \text{ for all } \quad Y \in \mathbb{H}_n.$$

We need a slight modification of this result. If

$$h(X) = \|\log(A^{-1/2}XA^{-1/2})\|_2^2,$$

then

$$\begin{aligned}&Dh(X)(Y) \\ &= 2\Big\langle (A^{-1/2}XA^{-1/2})^{-1} \log(A^{-1/2}XA^{-1/2}), A^{-1/2}YA^{-1/2} \Big\rangle\end{aligned} \tag{6.25}$$

for all $Y \in \mathbb{H}_n$.

6.3.3 Theorem

Let A_1, A_2, A_3 be any three elements of $\mathbb{P}_n$, and let

$$f(X) = \sum_{j=1}^{3} \delta_2^2(A_j, X). \tag{6.26}$$

Then the derivative of f at X is given by

$$Df(X)(Y) = 2\sum_{j=1}^{3} \left\langle X^{-1} \ \log{(XA_j^{-1})}, Y \right\rangle, \tag{6.27}$$

for all $Y \in \mathbb{H}_n$.

Proof. Using the relation (6.13) we have

$$f(X) = \sum_{j=1}^{3} \| \log \left(A_j^{-1/2} X A_j^{-1/2} \right) \|_2^2.$$

Using (6.25) we see that $Df(X)(Y)$ is a sum of three terms of the form

$$\begin{aligned} &2 \ \mathrm{tr} \left[A_j^{1/2} X^{-1} A_j^{1/2} \ \log \left(A_j^{-1/2} X A_j^{-1/2} \right) A_j^{-1/2} Y A_j^{-1/2} \right] \\ &= 2 \,\mathrm{tr} \left[X^{-1} A_j^{1/2} \ \log \left(A_j^{-1/2} X A_j^{-1/2} \right) A_j^{-1/2} Y \right] \\ &= 2 \,\mathrm{tr} \left[X^{-1} \ \log \left(X A_j^{-1} \right) Y \right]. \end{aligned}$$

Here we have used the similarity invariance of trace at the first step, and then the relation

$$S \ \log(T) S^{-1} = \log(STS^{-1})$$

at the second step. The latter is valid for all matrices T with no eigenvalues on the half-line $(-\infty, 0]$ and for all invertible matrices S, and follows from the usual functional calculus. This proves the theorem. ∎

6.3.4 Theorem

Let A_1, A_2, A_3 be three positive matrices and let $X_0 = G(A_1, A_2, A_3)$ be the point defined by (6.24). Then X_0 is the unique positive solution of the equation

$$\sum_{j=1}^{3} X^{-1} \log(XA_j^{-1}) = O. \tag{6.28}$$

Proof. The point X_0 is the unique minimum of the function (6.26), and hence, is characterised by the vanishing of the derivative (6.27) for all $Y \in \mathbb{H}_n$. But any matrix orthogonal to all Hermitian matrices is zero. Hence

$$\sum_{j=1}^{3} X_0^{-1} \log(X_0 A_j^{-1}) = O. \tag{6.29}$$

In other words X_0 satisfies the equation (6.28). ■

6.3.5 Exercise

Let A_1, A_2, A_3 be pairwise commuting positive matrices. Show that $G(A_1, A_2, A_3) = (A_1 A_2 A_3)^{1/3}$.

6.3.6 Exercise

Let X and A be positive matrices. Show that

$$X^{-1} \log(XA^{-1}) = X^{-1/2} \log\left(X^{1/2} A^{-1} X^{1/2}\right) X^{-1/2}. \tag{6.30}$$

(This shows that the matrices occurring in (6.29) are Hermitian.)

6.3.7 Exercise

Let $w = (w_1, w_2, w_3)$, where $w_j \geq 0$ and $\sum w_j = 1$. We say that w is a set of weights. Let

$$f_w(X) = \sum_{j=1}^{3} w_j \delta_2^2(A_j, X).$$

Show that f_w is strictly convex, and attains a minimum at a unique point.

Let $G_w(A_1, A_2, A_3)$ be the point where f_w attains its minimum. The special choice $w = (1/3, 1/3, 1/3)$ leads to $G(A_1, A_2, A_3)$.

6.3.8 Proposition

Each of the points $G_w(A_1, A_2, A_3)$ *lies in the closure of the convex hull* $\mathrm{conv}\,(\{A_1, A_2, A_3\})$.

Proof. Let $\mathcal{K}$ be the closure of $\mathrm{conv}\,(\{A_1, A_2, A_3\})$ and let π be the metric projection onto $\mathcal{K}$. Then by Theorem 6.2.7, $\delta_2^2(A_j, X) \geq \delta_2^2(A_j, \pi(X))$ for every $X \in \mathbb{P}_n$. Hence $f_w(X) \geq f_w(\pi(X))$ for all X. Thus the minimum value of $f_w(X)$ cannot be attained at a point outside $\mathcal{K}$. ■

Now we turn to another possible definition of the geometric mean of three matrices inspired by the characterisation of the centre of a triangle as the intersection of a sequence of nested triangles.

Given A_1, A_2, A_3 in $\mathbb{P}_n$ inductively construct a sequence of triples $\left\{A_1^{(m)}, A_2^{(m)}, A_3^{(m)}\right\}$ as follows. Set $A_1^{(0)} = A_1,\ A_2^{(0)} = A_2,\ A_3^{(0)} = A_3$, and let

$$A_1^{(m+1)} = A_1^{(m)} \# A_2^{(m)},\ A_2^{(m+1)} = A_2^{(m)} \# A_3^{(m)},\ A_3^{(m+1)} = A_3^{(m)} \# A_1^{(m)}. \tag{6.31}$$

6.3.9 Theorem

Let A_1, A_2, A_3 *be any three points in* $\mathbb{P}_n$, *and let* $\left\{A_1^{(m)}, A_2^{(m)}, A_3^{(m)}\right\}$ *be the sequence defined by (6.31). Then for any choice of* X_m *in* $\mathrm{conv}\left(\left\{A_1^{(m)}, A_2^{(m)}, A_3^{(m)}\right\}\right)$ *the sequence* $\{X_m\}$ *converges to a point* $X \in \mathrm{conv}\,(\{A_1, A_2, A_3\})$. *The point* X *does not depend on the choice of* X_m.

Proof. The diameter of a set $\mathcal{S}$ in $\mathbb{P}_n$ is defined as

$$\mathrm{diam}\,\mathcal{S} = \sup\{\delta_2(X, Y) : X, Y \in \mathcal{S}\}.$$

It is easy to see, using convexity of the metric δ_2, that if $\mathrm{diam}\,\mathcal{S} = M$, then $\mathrm{diam}\,(\mathrm{conv}\,(\mathcal{S})) = M$.

Let $\mathcal{K}_m = \mathrm{conv}\left(\left\{A_1^{(m)}, A_2^{(m)}, A_3^{(m)}\right\}\right)$. By (6.17), and what we said above, $\mathrm{diam}\,\mathcal{K}_m \leq 2^{-m} M_0$, where $M_0 = \mathrm{diam}\,\{A_1, A_2, A_3\}$. The sequence $\{\mathcal{K}_m\}$ is a decreasing sequence. Hence $\{X_m\}$ is Cauchy and converges to a limit X. Since X_m is in $\mathcal{K}_0$ for all m, the limit X is in the closure of $\mathcal{K}_0$. The limit is unique as any two such sequences can be interlaced. ■

6.3.10 A geometric mean of three matrices

Let $G^{\#}(A_1, A_2, A_3)$ be the limit point X whose existence has been proved in Theorem 6.3.9. This may be thought of as a geometric mean of A_1, A_2, A_3. From its construction it is clear that $G^{\#}$ is a symmetric continuous function of A_1, A_2, A_3. Since the geometric mean $A\#B$ of two matrices is monotone in A and B and is invariant under congruence transformations, these properties are inherited by $G^{\#}(A_1, A_2, A_3)$ as its construction involves successive two-variable means and limits.

Exercise Show that for a commuting triple A_1, A_2, A_3 of positive matrices $G^{\#}(A_1, A_2, A_3) = (A_1 A_2 A_3)^{1/3}$.

One may wonder whether $G^{\#}(A_1, A_2, A_3)$ is equal to the centre of mass $G(A_1, A_2, A_3)$. It turns out that this is not always the case. Thus we have here two different candidates for a geometric mean of three matrices. While $G^{\#}$ has all properties that we seek, it is not known whether G is monotone in its arguments. It does have all other desired properties.

6.4 RELATED INEQUALITIES

Some of the inequalities proved in Section 6.1 can be generalized from the special $\|\cdot\|_2$ norm to all Schatten $\|\cdot\|_p$ norms and to the larger class of unitarily invariant norms. These inequalities are very closely related to others proved in very different contexts like quantum statistical mechanics. This section is a brief indication of these connections.

Two results from earlier chapters provide the basis for our generalizations. In Exercise 2.7.12 we saw that for a positive matrix A

$$|||A \circ X||| \leq \max \; a_{ii} |||X|||$$

for every X and every unitarily invariant norm. In Section 5.2.9 we showed that for every choice of n positive numbers $\lambda_1, \ldots, \lambda_n$, the matrix

$$\left[\left[\frac{\sinh(\lambda_i - \lambda_j)}{\lambda_i - \lambda_j} \right]\right]$$

is positive. Using these we can easily prove the following generalized version of Proposition 6.1.2.

6.4.1 Proposition (Generalized IEMI)

For all H and K in $\mathbb{H}_n$ we have

$$|||e^{-H/2}De^{H}(K)e^{-H/2}||| \geq |||K||| \tag{6.32}$$

for every unitarily invariant norm.

In the definition (6.2) replace $\|\cdot\|_2$ by any unitarily invariant norm $|||\cdot|||$ and call the resulting length $L_{|||\cdot|||}$; i.e.,

$$L_{|||\cdot|||}(\gamma) = \int_a^b |||\gamma^{-1/2}(t)\gamma'(t)\gamma^{-1/2}(t)|||\; dt. \tag{6.33}$$

Since $|||X|||$ is a (symmetric gauge) function of the singular values of X, Lemma 6.1.1 carries over to $L_{|||\cdot|||}$. The analogue of (6.4),

$$\delta_{|||\cdot|||}(A,B) = \inf\left\{L_{|||\cdot|||}(\gamma) : \gamma \text{ is a path from } A \text{ to } B\right\}, \tag{6.34}$$

is a metric on $\mathbb{P}_n$ invariant under congruence transformations. The generalized IEMI leads to a generalized EMI. For all A, B in $\mathbb{P}_n$ we have

$$\delta_{|||\cdot|||}(A,B) \geq |||\log A - \log B|||, \tag{6.35}$$

or, in other words, for all H, K in $\mathbb{H}_n$

$$\delta_{|||\cdot|||}(e^H, e^K) \geq |||H - K|||. \tag{6.36}$$

Some care is needed while formulating statements about uniqueness of geodesics. Many unitarily invariant norms have the property that, in the metric they induce on $\mathbb{H}_n$, the straight line segment is the unique geodesic joining any two given points. If a norm $|||\cdot|||$ has this property, then the metric $\delta_{|||\cdot|||}$ on $\mathbb{P}_n$ inherits it. The Schatten p-norms have this property for $1 < p < \infty$, but not for $p = 1$ or ∞. With this proviso, statements made in Sections 6.1.5 and 6.1.6 can be proved in the more general setting. In particular, we have

$$\delta_{|||\cdot|||}(A,B) = |||\log A^{-1/2}BA^{-1/2}|||. \tag{6.37}$$

The geometric mean $A\#B$ defined by (4.10) is equidistant from A and B in each of the metrics $\delta_{|||\cdot|||}$. For certain metrics, such as the ones corresponding to Schatten p-norms for $1 < p < \infty$, this is the unique "metric midpoint" between A and B.

The parallelogram law and the semiparallelogram law, however, characterize a Hilbert space norm and the associated Riemannian metric. These are not valid for other metrics.

Now we can see the connection between these inequalities arising from geometry to others related to physics. Some facts about majorization and unitarily invariant norms are needed in the ensuing discussion. Let H, K be Hermitian matrices. From (6.36) and (6.37) we have

$$|||H+K||| \leq ||| \log(e^{H/2} e^K e^{H/2}) |||. \tag{6.38}$$

The exponential function is convex and monotonically increasing on $\mathbb{R}$. Such functions preserve weak majorization (Corollary II.3.4 in MA). Using this property we obtain from the inequality (6.38)

$$|||e^{H+K}||| \leq |||e^{H/2} e^K e^{H/2}|||. \tag{6.39}$$

Two special cases of this are well-known inequalities in physics. The special cases of the $\|\cdot\|_1$ and the $\|\cdot\|$ norms in (6.39) say

$$\operatorname{tr} e^{H+K} \leq \operatorname{tr} e^H e^K \tag{6.40}$$

and

$$\lambda_1(e^{H+K}) \leq \lambda_1(e^H e^K), \tag{6.41}$$

where $\lambda_1(X)$ is the largest eigenvalue of a matrix with real eigenvalues. The first of these is called the *Golden-Thompson inequality* and the second is called *Segal's inequality*.

The inequality (6.41) can be easily derived from the operator monotonicity of the logarithm function (Exercise 4.2.5 and Section 5.3.7). Let

$$\alpha = \lambda_1\left(e^H e^K\right) = \lambda_1\left(e^{K/2} e^H e^{K/2}\right).$$

Then

$$e^{K/2} e^H e^{K/2} \leq \alpha I,$$

and hence

$$e^H \leq \alpha e^{-K}.$$

Since log is an operator monotone function on $(0, \infty)$, it follows that

$$H \leq (\log \alpha) I - K.$$

Hence

$$H + K \leq (\log \alpha) I$$

and therefore

$$e^{H+K} \leq \alpha I.$$

This leads to (6.41).

More interrelations between various inequalities are given in the next section and in the notes at the end of the chapter.

6.5 SUPPLEMENTARY RESULTS AND EXERCISES

The crucial inequality (6.6) has a short alternate proof based on the inequality between the geometric and the logarithmic means. This relies on the following interesting formula for the derivative of the exponential map:

$$De^X(Y) = \int_0^1 e^{tX} Y e^{(1-t)X} dt. \tag{6.42}$$

This formula, attributed variously to Duhamel, Dyson, Feynman, and Schwinger, has an easy proof. Since

$$\frac{d}{dt}\left(e^{tX} e^{(1-t)Y}\right) = e^{tX}(X - Y)e^{(1-t)Y},$$

we have

$$e^X - e^Y = \int_0^1 e^{tX}(X - Y)e^{(1-t)Y} dt.$$

Hence

$$\lim_{h \to 0} \frac{e^{X+hY} - e^X}{h} = \int_0^1 e^{tX} Y e^{(1-t)X} dt.$$

This is exactly the statement (6.42).

Now let H and K be Hermitian matrices. Using the identity

$$K = e^{H/2}\left(e^{-H/2} K e^{-H/2}\right)e^{H/2}$$

and the first inequality in (5.34) we obtain

$$\begin{aligned} \|K\|_2 &\leq \left\| \int_0^1 e^{tH}\left(e^{-H/2} K e^{-H/2}\right) e^{(1-t)H} dt \right\|_2 \\ &= \left\| e^{-H/2}\left[\int_0^1 e^{tH} K e^{(1-t)H} dt\right] e^{-H/2} \right\|_2. \end{aligned}$$

The last integral is equal to $De^H(K)$. Hence,

$$\|K\|_2 \leq \|e^{-H/2} De^H(K) e^{-H/2}\|_2.$$

This is the IEMI (6.6).

The inequality (5.35) generalizes (5.34) to all unitarily invariant norms. So, exactly the same argument as above leads to a proof of (6.32) as well.

From the expression (6.14) it is clear that

$$\delta_2(A^{-1}, B^{-1}) = \delta_2(A, B), \tag{6.43}$$

for all $A, B \in \mathbb{P}_n$. Similarly, from (6.37) we see that

$$\delta_{|||\cdot|||}(A^{-1}, B^{-1}) = \delta_{|||\cdot|||}(A, B). \tag{6.44}$$

An important notion in geometry is that of a *Riemannian symmetric space.* By definition, this is a connected Riemannian manifold M for each point p of which there is an isometry σ_p of M with two properties:

(i) $\sigma_p(p) = p$, and

(ii) the derivative of σ_p at p is multiplication by -1.

The space $(\mathbb{P}_n, \delta_2)$ is a Riemannian symmetric space. We show this using the notation and some basic facts on matrix differential calculus from Section X.4 of MA. For each $A \in \mathbb{P}_n$ let σ_A be the map defined on $\mathbb{P}_n$ by

$$\sigma_A(X) = AX^{-1}A.$$

Clearly $\sigma_A(A) = A$. Let $\mathcal{I}(X) = X^{-1}$ be the inversion map. Then σ_A is the composite $\Gamma_A \cdot \mathcal{I}$. The derivative of $\mathcal{I}$ is given by $D\mathcal{I}(X)(Y) = -X^{-1}YX^{-1}$, while Γ_A being a linear map is equal to its own derivative. So, by the chain rule

$$\begin{aligned} D\sigma_A(A)(Y) &= D\Gamma_A\Big(\mathcal{I}(A)\Big) D\mathcal{I}(A)(Y) \\ &= A\Big(-A^{-1}YA^{-1}\Big)A = -Y. \end{aligned}$$

Thus $D\sigma_p(A)$ is multiplication by -1.

The Riemannian manifold $\mathbb{P}_n$ has *nonpositive curvature.* The EMI captures the essence of this fact. We explain this briefly.

Consider a triangle $\triangle(O, H, K)$ with vertices O, H, and K in $\mathbb{H}_n$. The image of this set under the exponential map is a "triangle"

$\triangle(I, e^H, e^K)$ in $\mathbb{P}_n$. By Proposition 6.1.5 the δ_2-lengths of the sides $[I, e^H]$ and $[I, e^K]$ are equal to the $\|\cdot\|_2$-lengths of the sides $[O, H]$ and $[O, K]$, respectively. By the EMI (6.8) the third side $[e^H, e^K]$ is longer than $[H, K]$. Keep the vertex O as a fixed pivot and move the sides $[O, H]$ and $[O, K]$ apart to get a triangle $\triangle(O, H', K')$ in $\mathbb{H}_n$ whose three sides now have the same lengths as the δ_2-lengths of the sides of $\triangle(I, e^H, e^K)$ in $\mathbb{P}_n$. Such a triangle is called a *comparison triangle* for $\triangle(I, e^H, e^K)$ and it is unique up to an isometry of $\mathbb{H}_n$. The fact that the comparison triangle in the Euclidean space $\mathbb{H}_n$ is "fatter" than the triangle $\triangle(I, e^H, e^K)$ is a characterization of a space of nonpositive curvature.

It may be instructive here to compare the situation with the space $\mathbb{U}_n$ consisting of unitary matrices. This is a compact manifold of *nonnegative curvature.* In this case the real vector space $i\mathbb{H}_n$ consisting of skew-Hermitian matrices is mapped by the exponential onto $\mathbb{U}_n$. The map is not injective in this case; it is a local diffeomorphism.

6.5.1 Exercise

Let H and K be any two skew-Hermitian matrices. Show that

$$\|De^H(K)\|_2 \le \|K\|_2. \tag{6.45}$$

[Hint: Follow the steps in the proof of Proposition 6.1.2. Now the λ_i are imaginary. So the hyperbolic function sinh occurring in the proof of Proposition 6.1.2 is replaced by the circular function sin. Alternately prove this using the formula (6.42). Observe that e^{tH} is unitary.]

As a consequence we have the opposite of the inequality (6.8) in this case: if A and B are sufficiently close in $\mathbb{U}_n$, then

$$\delta_2(A, B) \le \|\log A - \log B\|_2.$$

Thus the exponential map decreases distance locally. This fact captures the nonnegative curvature of $\mathbb{U}_n$.

Of late there has been interest in general *metric spaces of nonpositive curvature* (not necessarily Riemannian manifolds). An important consequence of the generalised EMI proved in Section 6.4 is that for every unitarily invariant norm the space $(\mathbb{P}_n, \delta_{|||\cdot|||})$ is a metric space of nonpositive curvature. These are examples of Finsler manifolds, where the metric arises from a non-Euclidean metric on the tangent space.

A metric space (X, d) is said to satisfy the *semiparallelogram law* if for any two points $a, b \in X$, there exists a point m such that

$$d^2(a,b) + 4d^2(m,c) \leq 2d^2(a,c) + 2d^2(b,c) \tag{6.46}$$

for all $c \in X$.

6.5.2 Exercise

Let (X, d) be a metric space with the semiparallelogram law. Show that the point m arising in the definition is unique and is the metric midpoint of a and b; i.e., m is the point at which $d(a,m) = d(b,m) = \frac{1}{2}d(a,b)$.

A complete metric space satisfying the semiparallelogram law is called a *Bruhat-Tits space.* We have shown that $(\mathbb{P}_n, \delta_2)$ is such a space. Those of our proofs that involved only completeness and the semiparallelogram law are valid for all Bruhat-Tits spaces. See, for example, Theorems 6.2.6 and 6.2.7.

In the next two exercises we point out more connections between classical matrix inequalities and geometric facts of this chapter. We use the notation of majorization and facts about unitarily invariant norms from MA, Chapters II and IV. The reader unfamiliar with these may skip this part.

6.5.3 Exercise

An inequality due to Gel'fand, Naimark, and Lidskii gives relations between eigenvalues of two positive matrices A and B and their product AB. This says

$$\log \lambda^{\downarrow}(A) + \log \lambda^{\uparrow}(B) \prec \log \lambda(AB) \prec \log \lambda^{\downarrow}(A) + \log \lambda^{\downarrow}(B). \tag{6.47}$$

See MA p. 73. Let A, B, and C be three positive matrices. Then

$$\begin{aligned}\lambda(A^{-1}C) &= \lambda\left(B^{1/2}A^{-1}CB^{-1/2}\right)\\ &= \lambda\left(B^{1/2}A^{-1}B^{1/2}B^{-1/2}CB^{-1/2}\right).\end{aligned}$$

So, by the second part of (6.47)

$$\begin{aligned}\log\, \lambda(A^{-1}C) &\prec \log\, \lambda^{\downarrow}\left(B^{1/2}A^{-1}B^{1/2}\right) + \log \lambda^{\downarrow}\left(B^{-1/2}CB^{-1/2}\right)\\ &= \log\, \lambda^{\downarrow}\left(A^{-1}B\right) + \log\, \lambda^{\downarrow}\left(B^{-1}C\right).\end{aligned}$$

Use this to show directly that $\delta_{|||\cdot|||}$ defined by (6.36) is a metric on $\mathbb{P}_n$.

6.5.4 Exercise

Let A and B be positive. Then for $0 \le t \le 1$ and $1 \le k \le n$ we have

$$\prod_{j=1}^{k} \lambda_j \left(B^{-t/2} A^t B^{-t/2}\right) \le \prod_{j=1}^{k} \lambda_j^t \left(B^{-1/2} A B^{-1/2}\right). \tag{6.48}$$

See MA p. 258. Take logarithms of both sides and use results on majorization to show that

$$|||\log\, B^{-t/2} A^t B^{-t/2}||| \le t\ \ |||\log\, B^{-1/2} A B^{-1/2}|||.$$

This may be rewritten as

$$\delta_{|||\cdot|||}(A^t, B^t) \le t\ \ \delta_{|||\cdot|||}(A, B), \quad 0 \le t \le 1.$$

Show that this implies that the metric $\delta_{|||\cdot|||}$ is convex.

In Section 4.5 we outlined a general procedure for constructing matrix means from scalar means. Two such means are germane to our present discussion. The function f in (4.69) corresponding to the logarithmic mean is

$$f(x) = \int_0^1 x^t dt.$$

So the logarithmic mean of two positive matrices A and B given by the formula (4.71) is

$$L(A, B) = A^{1/2} \int_0^1 \left(A^{-1/2} B A^{-1/2}\right)^t dt\ \ A^{1/2}.$$

In other words

$$L(A, B) = \int_0^1 \gamma(t) dt, \tag{6.49}$$

where $\gamma(t)$ is the geodesic segment joining A and B.

Likewise, for $0 \le t \le 1$ the Heinz mean

$$H_t(a, b) = \frac{a^t b^{1-t} + a^{1-t} b^t}{2} \tag{6.50}$$

leads to the function

$$f_t(x) = H_t(x, 1) = \frac{x^t + x^{1-t}}{2},$$

and then to the matrix Heinz mean

$$H_t(A,B) = \frac{\gamma(t) + \gamma(1-t)}{2}. \tag{6.51}$$

The following theorem shows that the geodesic $\gamma(t)$ has very intimate connections with the order relation on $\mathbb{P}_n$.

6.5.5 Theorem

For every α in $[0, 1/2]$ we have

$$\begin{aligned} A\#B &\leq \frac{1}{2\alpha}\int_{1/2-\alpha}^{1/2+\alpha} \gamma(t)dt \\ &\leq \int_0^1 \gamma(t)dt \\ &\leq \frac{1}{2\alpha}\left[\int_0^{\alpha} \gamma(t)dt + \int_{1-\alpha}^{1} \gamma(t)dt\right] \\ &\leq \frac{A+B}{2}. \end{aligned}$$

Proof. It is enough to prove the scalar versions of these inequalities as they are preserved in the transition to matrices by our construction. For fixed a and b, $H_t(a,b)$ is a convex function of t on $[0,1]$. It is symmetric about the point $t = 1/2$ at which it attains its minimum. Hence the quantity

$$\frac{1}{2\alpha}\int_{1/2-\alpha}^{1/2+\alpha} H_t(a,b)dt = \frac{1}{2\alpha}\int_{1/2-\alpha}^{1/2+\alpha} a^t b^{1-t} dt$$

is an increasing function of α for $0 \leq \alpha \leq 1/2$. Similarly,

$$\frac{1}{2\alpha}\left[\int_0^{\alpha} + \int_{1-\alpha}^{1} H_t(a,b)dt\right] = \frac{1}{2\alpha}\left[\int_0^{\alpha} + \int_{1-\alpha}^{1} a^t b^{1-t} dt\right]$$

is a decreasing function of α. These considerations show

$$\begin{aligned} \sqrt{ab} &\leq \frac{1}{2\alpha}\int_{1/2-\alpha}^{1/2+\alpha} a^t b^{1-t} dt \leq \int_0^1 a^t b^{1-t} dt \\ &\leq \frac{1}{2\alpha}\left[\int_0^{\alpha} + \int_{1-\alpha}^{1} a^t b^{1-t} dt\right] \leq \frac{a+b}{2}. \end{aligned}$$

The theorem follows from this. ∎

6.5.6 Exercise

Show that for $0 \le t \le 1$

$$\gamma(t) \le (1-t)A + tB. \tag{6.52}$$

[Hint: Show that for each $\lambda > 0$ we have $\lambda^t \le (1-t) + t\lambda$.]

6.5.7 Exercise

Let Φ be any positive linear map on $\mathbb{M}_n$. Then for all positive matrices A and B

$$\Phi\Big(L(A,B)\Big) \le L\Big(\Phi(A), \Phi(B)\Big).$$

[Hint: Use Theorem 4.1.5 (ii).]

6.5.8 Exercise

The aim of this exercise is to give a simple proof of the convergence argument needed to establish the existence of $G^{\#}(A_1, A_2, A_3)$ defined in Section 6.3.10.

(i) Assume that $A_1 \le A_2 \le A_3$. Then the sequences defined in (6.31) satisfy

$$A_1^{(m)} \le A_2^{(m)} \le A_3^{(m)} \quad \text{for all} \quad m.$$

The sequence $\{A_1^{(m)}\}$ is increasing and $\{A_3^{(m)}\}$ is decreasing. Hence the limits

$$L = \lim_{m\to\infty} A_1^{(m)} \quad \text{and} \quad U = \lim_{m\to\infty} A_3^{(m)}$$

exist. Show that $L = U$. Thus

$$\lim_{m\to\infty} A_1^{(m)} = \lim_{m\to\infty} A_2^{(m)} = \lim_{m\to\infty} A_3^{(m)}.$$

Call this limit $G^{\#}(A_1, A_2, A_3)$.

(ii) Now let A_1, A_2, A_3 be any three positive matrices. Choose positive numbers λ and μ such that

$$A_1 < \lambda A_2 < \mu A_3.$$

Let $(B_1, B_2, B_3) = (A_1, \lambda A_2, \mu A_3)$. Apply the special case (i) to get the limit $G^{\#}(B_1, B_2, B_3)$. The same recursion applied to the triple of numbers $(a_1, a_2, a_3) = (1, \lambda, \mu)$ gives

$$\lim_{m\to\infty} a_j^{(m)} = (\lambda\mu)^{1/3} \quad \text{for } j = 1, 2, 3.$$

Since

$$A_j^{(m)} = \frac{B_j^{(m)}}{a_j^{(m)}} \quad \text{for all } m = 1, 2, \ldots;\ j = 1, 2, 3,$$

it follows that the sequences $A_j^{(m)}$, $j = 1, 2, 3$, converge to the limit $G^{\#}(B_1, B_2, B_3)/(\lambda\mu)^{1/3}$.

6.5.9 Exercise

Show that the center of mass defined by (6.24) has the property

$$G(A_1, A_2, A_3)^{-1} = G(A_1^{-1}, A_2^{-1}, A_3^{-1})$$

for all positive matrices A_1, A_2, A_3. Show that $G^{\#}$ also satisfies this relation.

6.6 NOTES AND REFERENCES

Much of the material in Sections 6.1 and 6.2 consists of standard topics in Riemannian geometry. The arrangement of topics, the emphasis, and some proofs are perhaps eccentric. Our view is directed toward applications in matrix analysis, and the treatment may provide a quick introduction to some of the concepts. The entire chapter is based on R. Bhatia and J. A. R. Holbrook, *Riemannian geometry and matrix geometric means,* Linear Algebra Appl., 413 (2006) 594–618.

Two books on Riemannian geometry that we recommend are M. Berger, *A Panoramic View of Riemannian Geometry,* Springer, 2003, and S. Lang, *Fundamentals of Differential Geometry,* Springer, 1999. Closely related to our discussion is M. Bridson and A. Haefliger, *Metric Spaces of Non-positive Curvature,* Springer, 1999. Most of the texts on geometry emphasize group structures and seem to downplay the role of the matrices that constitute these groups. Lang's text is exceptional in this respect. The book A. Terras, *Harmonic Analysis on Symmetric Spaces and Applications II,* Springer, 1988, devotes a long chapter to the space $\mathbb{P}_n$.

The proof of Proposition 6.1.2 is close to the treatment in Lang's book. (Lang says he follows "Mostow's very elegant exposition of Cartan's work.") The linear algebra in our proof looks neater because a part of the work has been done earlier in proving the Daleckii-Krein formula (2.40) for the derivative. The second proof given at the beginning of Section 6.5 is shorter and more elementary. This is taken from R. Bhatia, *On the exponential metric increasing property,* Linear Algebra Appl., 375 (2003) 211–220.

Explicit formulas like (6.11) describing geodesics are generally not emphasized in geometry texts. This expression has been used often in connection with means. With the notation $A\#_tB$ this is called the *t-power mean.* See the comprehensive survey F. Hiai, *Log-majorizations and norm inequalities for exponential operators,* Banach Center Publications Vol. 38, pp. 119–181.

The role of the semiparallelogram law is highlighted in Chapter XI of Lang's book. A historical note on page 313 of this book places it in context. To a reader oriented towards analysis in general, and inequalities in particular, this is especially attractive. The expository article by J. D. Lawson and Y. Lim, *The geometric mean, matrices, metrics and more,* Am. Math. Monthly, 108 (2001) 797–812, draws special attention to the geometry behind the geometric mean.

Problems related to convexity in differentiable manifolds are generally difficult. According to Note 6.1.3.1 on page 231 of Berger's book the problem of identifying the convex hull of three points in a Riemannian manifold of dimension 3 or more is still unsolved. It is not even known whether this set is closed. This problem is reflected in some of our difficulties in Section 6.3.

Berger attributes to E. Cartan, *Groupes simples clos et ouverts et géometrie Riemannienne,* J. Math. Pures Appl., 8 (1929) 1–33, the introduction of the idea of center of mass in Riemannian geometry. Cartan showed that in a complete manifold of nonpositive curvature (such as $\mathbb{P}_n$) every compact set has a unique center of mass. He used this to prove his fundamental theorem that says any two compact maximal subgroups of a semisimple Lie group are always conjugate.

The idea of using the center of mass to define a geometric mean of three positive matrices occurs in the paper of Bhatia and Holbrook cited earlier and in M. Moakher, *A differential geometric approach to the geometric mean of symmetric positive-definite matrices,* SIAM J. Matrix Anal. Appl., 26 (2005) 735–747. This paper contains many interesting ideas. In particular, Theorem 6.3.4 occurs here. Applications to problems of elasticity are discussed in M. Moakher, *On the averaging of symmetric positive-definite tensors,* preprint (2005).

The manifold $\mathbb{P}_n$ is the most studied example of a manifold of nonpositive curvature. However, one of its basic features—order—seems not to have received any attention. Our discussion of the center of mass and Theorem 6.5.5 show that order properties and geometric properties are strongly interlinked. A study of these properties should lead to a better understanding of this manifold.

The mean $G^{\#}(A_1, A_2, A_3)$ was introduced in T. Ando, C.-K Li, and R. Mathias, *Geometric Means,* Linear Algebra Appl., 385 (2004) 305–334. Many of its properties are derived in this paper which also contains a detailed survey of related matters. The connection with Riemannian geometry was made in the Bhatia-Holbrook paper cited earlier. That $G^{\#}$ and the center of mass may be different, is a conclusion made on the basis of computer-assisted numerical calculations reported in Bhatia-Holbrook. A better theoretical understanding is yet to be found.

As explained in Section 6.5 the EMI reflects the fact that $\mathbb{P}_n$ has nonpositive curvature. Inequalities of this type are called CAT(0) inequalities; the initials C, A, T are in honour of E. Cartan, A. D. Alexandrov, and A. Toponogov, respectively. These ideas have been given prominence in the work of M. Gromov. See the book W. Ballmann, M. Gromov, and V. Schroeder, *Manifolds of Nonpositive Curvature,* Birkhäuser, 1985, and the book by Bridson and Haefliger cited earlier. A concept of curvature for metric spaces (not necessarily Riemannian manifolds) is defined and studied in the latter. The generalised EMI proved in Section 6.4 shows that the space $\mathbb{P}_n$ with the metric $\delta_{|||\cdot|||}$ is a metric space (a Finsler manifold) of nonpositive curvature.

Segal's inequality was proved in I. Segal, *Notes towards the construction of nonlinear relativistic quantum fields III,* Bull. Am. Math. Soc., 75 (1969) 1390–1395. The simple proof given in Section 6.4 is borrowed from B. Simon, *Trace Ideals and Their Applications,* Second Edition, American Math. Society, 2005. The Golden-Thompson inequality is due to S. Golden, *Lower bounds for the Helmholtz function,* Phys. Rev. B, 137 (1965) 1127–1128, and C. J. Thompson, *Inequality with applications in statistical mechanics,* J. Math. Phys., 6 (1965) 1812–1813. Stronger versions and generalizations to other settings (like Lie groups) have been proved. Complementary inequalities have been proved by F. Hiai and D. Petz, *The Golden-Thompson trace inequality is complemented,* Linear Algebra Appl., 181 (1993) 153–185, and by T. Ando and F. Hiai, *Log majorization and complementary Golden-Thompson type inequalities,* ibid., 197/198 (1994) 113–131. These papers are especially interesting in our context as they involve the means $A\#_t B$ in the formulation and the proofs of several results. The connection between means, geodesics, and inequalities has been explored in several interesting papers by G. Corach and coauthors. Illustrative of this work and especially close to our discussion are the two papers by G. Corach, H. Porta and L. Recht, *Geodesics and operator means in the space of positive operators,* Int. J. Math., 4 (1993) 193–202, and *Convexity of the geodesic distance on spaces of positive operators,* Illinois J. Math., 38 (1994) 87–94.

The logarithmic mean $L(A, B)$ has not been studied before. The definition (6.49) raises interesting questions both for matrix theory and for geometry. In differential geometry it is common to integrate (real) functions along curves. Here we have the integral of the curve itself. Theorem 6.5.5 relates this object to other means, and includes the operator analogue of the inequality between the geometric, logarithmic, and arithmetic means. The norm version of this inequality appears as Proposition 3.2 in F. Hiai and H. Kosaki, *Means for matrices and comparison of their norms,* Indiana Univ. Math. J., 48 (1999) 899–936. Exercise 6.5.8 is based on the paper D. Petz and R. Temesi, *Means of positive numbers and matrices,* SIAM J. Matrix Anal. Appl., 27 (2005) 712–720.

Bibliography

[1] N. I. Akhiezer and I. M. Glazman, *Theory of Linear Operators in Hilbert Space,* Dover, 1993 (reprint of original edition).

[2] W. N. Anderson and R. J. Duffin, *Series and parallel addition of matrices,* J. Math. Anal. Appl., 26 (1969) 576–594.

[3] W. N. Anderson and G. E. Trapp, *Matrix operations induced by electrical network connections—a survey,* in *Constructive Approaches to Mathematical Models,* Academic Press, 1979, pp. 53–73.

[4] W. N. Anderson and G. E. Trapp, *Operator means and electrical networks,* in Proc. 1980 IEEE International Symposium on Circuits and Systems.

[5] T. Ando, *Structure of operators with numerical radius one,* Acta Sci. Math. (Szeged), 34 (1973) 11–15.

[6] T. Ando, *Topics on Operator Inequalities,* Hokkaido University, Sapporo, 1978.

[7] T. Ando, *Concavity of certain maps on positive definite matrices and applications to Hadamard products,* Linear Algebra Appl., 26 (1979) 203–241.

[8] T. Ando, *On some operator inequalities,* Math. Ann., 279 (1987) 157–159.

[9] T. Ando, *Operator-Theoretic Methods for Matrix Inequalities,* Sapporo, 1998.

[10] T. Ando and M.-D. Choi, *Non-linear completely positive maps,* in *Aspects of Positivity in Functional Analysis,* North Holland Mathematical Studies Vol. 122, 1986, pp.3–13.

[11] T. Ando and F. Hiai, *Log majorization and complementary Golden-Thompson type inequalities,* Linear Algebra Appl., 197/198 (1994) 113–131.

[12] T. Ando, C.-K Li, and R. Mathias, *Geometric Means,* Linear Algebra Appl., 385 (2004) 305–334.

[13] T. Ando and K. Okubo, *Induced norms of the Schur multiplier operator,* Linear Algebra Appl., 147 (1991) 181–199.

[14] T. Ando and X. Zhan, *Norm inequalities related to operator monotone functions,* Math. Ann., 315 (1999) 771–780.

[15] E. Andruchow, G. Corach, and D. Stojanoff, *Geometric operator inequalities,* Linear Algebra Appl., 258 (1997) 295–310.

[16] W. B. Arveson, *Subalgebras of* C^**-algebras, I, II,* Acta Math. 123 (1969) 141–224, and 128 (1972) 271–308.

[17] W. B. Arveson, *Nonlinear states on* C^**-algebras,* in *Operator Algebras and Mathematical Physics,* Contemporary Mathematics Vol. 62, American Math. Society, 1987, pp 283–343.

[18] W. Ballmann, M. Gromov, and V. Schroeder, *Manifolds of Nonpositive Curvature,* Birkhäuser, 1985.

[19] R. B. Bapat and T.E.S. Raghavan, *Nonnegative Matrices and Applications,* Cambridge University Press, 1997.

[20] C. Berg, J.P.R. Christensen, and P. Ressel, *Harmonic Analysis on Semigroups,* Springer, 1984.

[21] M. Berger, *A Panoramic View of Riemannian Geometry,* Springer, 2003.

[22] D. Bessis, P. Moussa, and M. Villani, *Monotonic converging variational approximations to the functional integrals in quantum statistical mechanics,* J. Math. Phys., 16 (1975) 2318–2325.

[23] K. V. Bhagwat and R. Subramanian, *Inequalities between means of positive operators,* Math. Proc. Cambridge Philos. Soc., 83 (1978) 393–401.

[24] R. Bhatia, *Variation of symmetric tensor powers and permanents,* Linear Algebra Appl., 62 (1984) 269–276.

[25] R. Bhatia, *First and second order perturbation bounds for the operator absolute value,* Linear Algebra Appl., 208/209 (1994) 367–376.

[26] R. Bhatia, *Matrix Analysis,* Springer, 1997.

[27] R. Bhatia, *Pinching, trimming, truncating and averaging of matrices,* Am. Math. Monthly, 107 (2000) 602–608.

[28] R. Bhatia, *Partial traces and entropy inequalities,* Linear Algebra Appl., 375 (2003) 211–220.

[29] R. Bhatia, *On the exponential metric increasing property,* Linear Algebra Appl., 375 (2003) 211–220.

[30] R. Bhatia, *Infinitely divisible matrices,* Am. Math. Monthly, 113 (2006) 221–235.

[31] R. Bhatia, *Interpolating the arithmetic-geometric mean inequality and its operator version,* Linear Algebra Appl., 413 (2006) 355–363.

[32] R. Bhatia, M.-D. Choi, and C. Davis, *Comparing a matrix to its off-diagonal part,* Oper. Theory: Adv. Appl., 40 (1989) 151–164.

[33] R. Bhatia and C. Davis, *More matrix forms of the arithmetic-geometric mean inequality,* SIAM J. Matrix Anal. Appl., 14 (1993) 132–136.

[34] R. Bhatia and C. Davis, *A better bound on the variance,* Am. Math. Monthly, 107 (2000) 602–608.

[35] R. Bhatia and C. Davis, *More operator versions of the Schwarz inequality,* Commun. Math. Phys., 215 (2000) 239–244.

[36] R. Bhatia and J. A. Dias da Silva, *Variation of induced linear operators,* Linear Algebra Appl., 341 (2002) 391–402.

[37] R. Bhatia and D. Drissi, *Generalised Lyapunov equations and positive definite functions,* SIAM J. Matrix Anal. Appl., 27 (2005) 103–114.

[38] R. Bhatia and L. Elsner, *Positive linear maps and the Lyapunov equation,* Oper. Theory: Adv. Appl., 130 (2001) 107–120.

[39] R. Bhatia and S. Friedland, *Variation of Grassman powers and spectra,* Linear Algebra Appl., 40 (1981) 1–18.

[40] R. Bhatia and J.A.R. Holbrook, *Fréchet derivatives of the power function,* Indiana Univ. Math. J., 49 (2000) 1155–1173.

[41] R. Bhatia and J.A.R. Holbrook, *Noncommutative geometric means,* Math. Intelligencer, 28 (2006) 32–39.

[42] R. Bhatia and J.A.R. Holbrook, *Riemannian geometry and matrix geometric means,* Linear Algebra Appl., 413 (2006) 594–618.

[43] R. Bhatia and F. Kittaneh, *On the singular values of a product of operators,* SIAM J. Matrix Anal. Appl., 11 (1990) 272–277.

[44] R. Bhatia and F. Kittaneh, *Norm inequalities for positive operators,* Lett. Math. Phys. 43 (1998) 225–231.

[45] R. Bhatia and F. Kittaneh, *Notes on matrix arithmetic-geometric mean inequalities,* Linear Algebra Appl., 308 (2000) 203–211.

[46] R. Bhatia and H. Kosaki, *Mean matrices and infinite divisibility,* Linear Algebra Appl., to appear.

[47] R. Bhatia and K. R. Parthasarathy, *Positive definite functions and operator inequalities,* Bull. London Math. Soc., 32 (2000) 214–228.

[48] R. Bhatia, D. Singh, and K. B. Sinha, *Differentiation of operator functions and perturbation bounds,* Commun. Math. Phys., 191 (1998) 603–611.

[49] R. Bhatia and K. B. Sinha, *Variation of real powers of positive operators,* Indiana Univ. Math. J., 43 (1994) 913–925.

[50] R. Bhatia and K. B. Sinha, *Derivations, derivatives and chain rules,* Linear Algebra Appl., 302/303 (1999) 231–244.

[51] M. Sh. Birman and M. Z. Solomyak, *Double Stieltjes operator integrals* (English translation), in *Topics in Mathematical Physics Vol. 1,* Consultant Bureau, New York, 1967.

[52] S. Bochner, *Vorlesungen über Fouriersche Integrale,* Akademie-Verlag, Berlin, 1932.

[53] L. Breiman, *Probability,* Addison-Wesley, 1968.

[54] M. Bridson and A. Haefliger, *Metric Spaces of Non-Positive Curvature,* Springer, 1999.

[55] H. J. Carlin and G. A. Noble, *Circuit properties of coupled dispersive lines with applications to wave guide modelling,* in *Proc. Network and Signal Theory,* J. K. Skwirzynski and J. O. Scanlan, eds., Peter Pergrinus, 1973, pp. 258–269.

[56] E. Cartan, *Groupes simples clos et ouverts et géometrie Riemannienne,* J. Math. Pures Appl., 8 (1929) 1–33.

[57] M.-D. Choi, *Positive linear maps on C^*-algebras,* Canadian J. Math., 24 (1972) 520–529.

[58] M.-D. Choi, *A Schwarz inequality for positive linear maps on C^*-algebras,* Illinois J. Math., 18 (1974) 565–574.

[59] M.-D. Choi, *Completely positive linear maps on complex matrices,* Linear Algebra Appl., 10 (1975) 285–290.

[60] M.-D. Choi, *Some assorted inequalities for positive linear maps on C^*-algebras,* J. Oper. Theory, 4 (1980) 271–285.

[61] M.-D. Choi, *Positive linear maps,* in *Operator Algebras and Applications, Part 2,* R. Kadison ed., American Math. Society, 1982.

[62] G. Corach, H. Porta, and L. Recht, *Geodesics and operator means in the space of positive operators,* Int. J. Math., 4 (1993) 193–202.

[63] G. Corach, H. Porta and L. Recht, *Convexity of the geodesic distance on spaces of positive operators,* Illinois J. Math., 38 (1994) 87–94.

[64] R. Cuppens, *Decomposition of Multivariate Probabilities,* Academic Press, 1975.

[65] Ju. L. Daleckii and S. G. Krein, *Formulas of differentiation according to a parameter of functions of Hermitian operators,* Dokl. Akad. Nauk SSSR, 76 (1951) 13–16.

[66] K. R. Davidson and J.A.R. Holbrook, *Numerical radii of zero-one matrices,* Michigan Math. J., 35 (1988) 261–267.

[67] C. Davis, *A Schwarz inequality for convex operator functions,* Proc. Am. Math. Soc., 8 (1957) 42–44.

[68] C. Davis, *Notions generalizing convexity for functions defined on spaces of matrices,* in *Proc. Symposia Pure Math., Vol. VII, Convexity,* American Math. Soc., 1963.

[69] J. De Pillis, *Transformations on partitioned matrices,* Duke Math. J., 36 (1969) 511–515.

[70] D. Drissi, *Sharp inequalities for some operator means,* SIAM J. Matrix Anal. Appl., to appear.

[71] S. W. Drury, S. Liu, C.-Y. Lu, S. Puntanen, and G.P.H. Styan, *Some comments on several matrix inequalities with applications to canonical correlations: historical background and recent developments,* Sankhyā, Series A, 64 (2002) 453–507.

[72] E. G. Effros and Z.-J. Ruan, *Operator Spaces,* Oxford University Press, 2000.

[73] W. G. Faris, *Review of S. J. Gustafson and I. M. Sigal, Mathematical Concepts of Quantum Mechanics,* SIAM Rev., 47 (2005) 379–380.

[74] M. Fiedler and V. Pták, *A new positive definite geometric mean of two positive definite matrices,* Linear Algebra Appl., 251 (1997) 1–20.

[75] T. Furuta, *$A \geq B \geq O$ assures $(B^r A^p B^r)^{1/q} \geq B^{(p+2r)/q}$ for $r \geq 0, p \geq 0, q \geq 1$ with $(1+2r)q \geq p+2r$*, Proc. Am. Math. Soc. 101 (1987) 85–88.

[76] F. Gantmacher, *Matrix Theory,* Chelsea, 1977.

[77] B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables,* Addison-Wesley, 1954.

[78] R. Goldberg, *Fourier Transforms,* Cambridge University Press, 1961.

[79] S. Golden, *Lower bounds for the Helmholtz function,* Phys. Rev. B, 137 (1965) 1127–1128.

[80] G. H. Golub and C. F. Van Loan, *Matrix Computations,* Third Edition, Johns Hopkins University Press, 1996.

[81] U. Haagerup, *Decomposition of completely bounded maps on operator algebras,* unpublished report.

[82] P. R. Halmos, *A Hilbert Space Problem Book,* Second Edition, Springer, 1982.

[83] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities,* Second Edition, Cambridge University Press, 1952.

[84] E. Haynsworth, *Determination of the inertia of a partitioned Hermitian matrix,* Linear Algebra Appl., 1 (1968) 73–81.

[85] E. Heinz, *Beiträge zur Störungstheorie der Spektralzerlegung,* Math. Ann., 123 (1951) 415–438.

[86] C. S. Herz, *Fonctions opérant sur les fonctions définies-positives,* Ann. Inst. Fourier (Grenoble), 13 (1963) 161–180.

[87] F. Hiai, *Log-majorizations and norm inequalities for exponential operators,* Banach Center Publications Vol. 38, pp. 119–181.

[88] F. Hiai and H. Kosaki, *Means for matrices and comparison of their norms,* Indiana Univ. Math. J., 48 (1999) 899–936.

[89] F. Hiai and H. Kosaki, *Comparison of various means for operators,* J. Funct. Anal., 163 (1999) 300–323.

[90] F. Hiai and H. Kosaki, *Means of Hilbert Space Operators,* Lecture Notes in Mathematics Vol. 1820, Springer, 2003.

[91] F. Hiai and D. Petz, *The Golden-Thompson trace inequality is complemented,* Linear Algebra Appl., 181 (1993) 153–185.

[92] N. J. Higham, *Accuracy and Stability of Numerical Algorithms,* Second Edition, SIAM, 2002.

[93] R. A. Horn, *On boundary values of a schlicht mapping,* Proc. Am. Math. Soc., 18 (1967) 782–787.

[94] R. A. Horn, *The Hadamard product,* in *Matrix Theory and Applications,* C. R. Johnson, ed., American Math. Society, 1990.

[95] R. A. Horn, *Norm bounds for Hadamard products and the arithmetic-geometric mean inequality for unitarily invariant norms,* Linear Algebra Appl., 223/224 (1995) 355–361.

[96] R. A. Horn and C. R. Johnson, *Matrix Analysis,* Cambridge University Press, 1985.

[97] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis,* Cambridge University Press, 1991.

[98] D. K. Jocić, *Cauchy-Schwarz and means inequalities for elementary operators into norm ideals,* Proc. Am. Math. Soc., 126 (1998) 2705–2711.

[99] D. K. Jocić, *Cauchy-Schwarz norm inequalities for weak***-integrals of operator valued functions,* J. Funct. Anal., 218 (2005) 318–346.

[100] C. R. Johnson, *Matrix completion problems: a survey,* Proc., Symposia in Applied Mathematics Vol. 40, American Math. Society, 1990.

[101] C. R. Johnson and C. J. Hillar, *Eigenvalues of words in two positive definite letters,* SIAM J. Matrix Anal. Appl., 23 (2002) 916–928.

[102] Z. J. Jurek and M. Yor, *Self decomposable laws associated with hyperbolic functions,* Probability and Mathematical Statistics, 24(2004), 181–190.

[103] R. Kadison, *A generalized Schwarz inequality and algebraic invariants for operator algebras,* Ann. of Math., 56 (1952) 494–503.

[104] H. Kosaki, *Arithmetic-geometric mean and related inequalities for operators,* J. Funct. Anal., 15 (1998) 429–451.

[105] H. Kosaki, *On infinite divisibility of positive definite functions,* preprint, 2006.

[106] K. Kraus, *General state changes in quantum theory,* Ann. Phys., 64 (1971) 311–335.

[107] K. Kraus, *States, Effects, and Operations: Fundamental Notions of Quantum Theory,* Lecture Notes in Physics Vol. 190, Springer, 1983.

[108] F. Kubo and T. Ando, *Means of positive linear operators,* Math. Ann., 246 (1980) 205–224.

[109] A. Kurosh, *Higher Algebra,* Mir Publishers, 1972.

[110] M. K. Kwong, *On the definiteness of the solutions of certain matrix equations,* Linear Algebra Appl., 108 (1988) 177–197.

[111] M. K. Kwong, *Some results on matrix monotone functions,* Linear Algebra Appl., 118 (1989) 129–153.

[112] P. Lancaster and L. Rodman, *The Algebraic Riccati Equation,* Oxford University Press, 1995.

[113] P. Lancaster and M. Tismenetsky, *The Theory of Matrices,* Second Edition, Academic Press, 1985.

[114] S. Lang, *Fundamentals of Differential Geometry,* Springer, 1999.

[115] J. D. Lawson and Y. Lim, *The geometric mean, matrices, metrics and more,* Am. Math. Monthly, 108 (2001) 797–812.

[116] P. Lax, *Linear Algebra,* John Wiley, 1997.

[117] P. Lévy, *Théorie de l' Addition des Variables Aléatoires,* Gauthier-Villars, 1937.

[118] E. H. Lieb, *Convex trace functions and the Wigner-Yanase-Dyson conjecture,* Adv. Math., 11 (1973) 267–288.

[119] E. H. Lieb, *Inequalities, Selecta of Elliot H. Lieb,* M. Loss and M. B. Ruskai eds., Springer, 2002.

[120] E. H. Lieb and M. B. Ruskai, *A fundamental property of quantum-mechanical entropy,* Phys. Rev. Lett., 30 (1973) 434–436.

[121] E. H. Lieb and M. B. Ruskai, *Proof of the strong subadditivity of quantum-mechanical entropy,* J. Math. Phys., 14 (1973) 1938–1941.

[122] E. H. Lieb and M. B. Ruskai, *Some operator inequalities of the Schwarz type,* Adv. Math., 12 (1974) 269–273.

[123] E. H. Lieb and R. Seiringer, *Equivalent forms of the Bessis-Moussa-Villani conjecture,* J. Stat. Phys., 115 (2004) 185–190.

[124] E. H. Lieb and J. Yngvason, *A guide to entropy and the second law of thermodynamics,* Notices Am. Math. Soc., 45 (1998) 571–581.

[125] E. H. Lieb and J. Yngvason, *The mathematical structure of the second law of thermodynamics,* in *Current Developments in Mathematics, 2001,* International Press, 2002.

[126] G. Lindblad, *Entropy, information and quantum measurements,* Commun. Math. Phys., 33 (1973) 305–322.

[127] G. Lindblad, *Expectations and entropy inequalities for finite quantum systems,* Commun. Math. Phys., 39 (1974) 111–119.

[128] G. Lindblad, *Completely positive maps and entropy inequalities,* Commun. Math. Phys., 40 (1975) 147–151.

[129] M. Loeve, *Probability Theory*, Van Nostrand, 1963.

[130] K. Löwner, *Über monotone Matrixfunctionen,* Math. Z., 38 (1934) 177–216.

[131] E. Lukacs, *Characteristic Functions,* Griffin, 1960.

[132] M. Marcus and W. Watkins, *Partitioned Hermitian matrices,* Duke Math. J., 38 (1971) 237–249.

[133] R. Mathias, *An arithmetic-geometric mean inequality involving Hadamard products,* Linear Algebra Appl., 184 (1993) 71–78.

[134] R. Mathias, *A note on: "More operator versions of the Schwarz inequality,"* Positivity, 8 (2004) 85–87.

[135] W. H. McAdams, *Heat Transmission,* Third Edition, McGraw-Hill, 1954.

[136] R. Merris, *Trace functions I,* Duke Math. J., 38 (1971) 527–530.

[137] P. A. Meyer, *Quantum Probability for Probabilists,* Lecture Notes in Mathematics Vol. 1538, Springer, 1993.

[138] M. Moakher, *A differential geometric approach to the geometric mean of symmetric positive-definite matrices,* SIAM J. Matrix Anal. Appl., 26 (2005) 735–747.

[139] M. Moakher, *On the averaging of symmetric positive-definite tensors,* preprint (2005).

[140] P. Moussa, *On the representation of* $Tr\left(e^{(A-\lambda B)}\right)$ *as a Laplace transform,* Rev. Math. Phys., 12 (2000) 621–655.

[141] V. Müller, *The numerical radius of a commuting product,* Michigan Math. J., 35 (1988) 255–260.

[142] J. von Neumann, *Thermodynamik quantenmechanischer Gesamtheiten,* Göttingen Nachr., 1927, pp. 273–291.

[143] J. von Neumann, *Mathematical Foundations of Quantum Mechanics,* Princeton University Press, 1955.

[144] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information,* Cambridge University Press, 2000.

[145] M. Ohya and D. Petz, *Quantum Entropy and Its Use,* Springer, 1993.

[146] K. R. Parthasarathy, *An Introduction to Quantum Stochastic Calculus,* Birkhäuser, 1992.

[147] K. R. Parthasarathy and K. Schmidt, *Positive Definite Kernels, Continuous Tensor Products and Central Limit Theorems of Probability Theory,* Lecture Notes in Mathematics Vol. 272, Springer, 1972.

[148] V. Paulsen, *Completely Bounded Maps and Dilations,* Longman, 1986.

[149] V. Paulsen, *Completely Bounded Maps and Operator Algebras,* Cambridge University Press, 2002.

[150] V. I. Paulsen, S. C. Power, and R. R. Smith, *Schur products and matrix completions,* J. Funct. Anal., 85 (1989) 151–178.

[151] D. Petz and R. Temesi, *Means of positive numbers and matrices,* SIAM J. Matrix Anal. Appl., 27 (2005) 712–720.

[152] G. Pisier, *Introduction to Operator Space Theory,* Cambridge University Press, 2003.

[153] G. Pólya, *Remarks on characteristic functions,* Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, 1949, pp. 115–123.

[154] W. Pusz and S. L. Woronowicz, *Functional calculus for sesquilinear forms and the purification map,* Rep. Math. Phys., 8 (1975) 159–170.

[155] M. Reed and B. Simon, *Methods of Modern Mathematical Physics, Vols. I, II,* Academic Press, 1972, 1975.

[156] B. Russo and H. A. Dye, *A note on unitary operators in C^*-algebras,* Duke Math. J., 33 (1966) 413–416.

[157] Z. Sasvári, *Positive Definite and Definitizable Functions,* Akademie-Verlag, Berlin, 1994.

[158] I. Schur, *Bemerkungen zur Theorie der beschränkten Bilinearformen mit unendlich vielen Veränderlichen,* J. Reine Angew. Math., 140 (1911), 1-28.

[159] I. Schur, *Über Potenzreihen die im Innern des Einheitskreises beschränkt sind* [I], J. Reine Angew. Math., 147 (1917) 205–232.

[160] L. Schwartz, *Théorie des Distributions,* Hermann, 1954.

[161] I. Segal, *Notes towards the construction of nonlinear relativistic quantum fields III,* Bull. Am. Math. Soc., 75 (1969) 1390–1395.

[162] C. Shannon, *Mathematical theory of communication,* Bell Sys. Tech. J., 27 (1948) 379–423.

[163] P. W. Shor, *Equivalence of additivity questions in quantum information theory,* Commun. Math. Phys., 246 (2004) 453–472.

[164] B. Simon, *Trace Ideals and Their Applications,* Second Edition, American Mathematical Society, 2005.

[165] J. M. Steele, *The Cauchy-Schwarz Master Class,* Math. Association of America, 2004.

[166] J. Stewart, *Positive definite functions and generalizations, an historical survey,* Rocky Mountain J. Math., 6 (1976) 409–434.

[167] W. F. Stinespring, *Positive functions on C^*-algebras,* Proc. Am. Math. Soc., 6 (1955) 211–216.

[168] E. Størmer, *Positive linear maps of operator algebras,* Acta Math., 110 (1963) 233–278.

[169] E. Størmer, *Positive linear maps of C^*-algebras,* in *Foundations of Quantum Mechanics and Ordered Linear Spaces,* Lecture Notes in Physics Vol. 29, Springer, 1974, pp. 85–106.

[170] V. S. Sunder, *A noncommutative analogue of* $|DX^k| = |kX^{k-1}|$, Linear Algebra Appl., 44 (1982) 87–95.

[171] A. Terras, *Harmonic Analysis on Symmetric Spaces and Applications II,* Springer, 1988.

[172] C. J. Thompson, *Inequality with applications in statistical mechanics,* J. Math. Phys., 6 (1965) 1812–1813.

[173] S.-G. Wang and W.-C. Ip, *A matrix version of the Wielandt inequality and its applications,* Linear Algebra Appl., 296 (1999) 171–181.

[174] A. Wehrl, *General properties of entropy,* Rev. Mod. Phys., 50 (1978) 221–260.

[175] A. Wehrl, *The many facets of entropy,* Rep. Math. Phys., 30 (1991) 119–129.

[176] S. L. Woronowicz, *Positive maps of low dimensional matrix algebras,* Rep. Math. Phys., 10 (1976) 165–183.

[177] X. Zhan, *Matrix Inequalities*, Lecture Notes in Mathematics Vol. 1790, Springer, 2002.

[178] F. Zhang, *Matrix Theory: Basic Results and Techniques,* Springer, 1999.

[179] F. Zhang (editor), *The Schur Complement and Its Applications,* Springer, 2005.

Index

Notation