INEQUALITIES

BY

EDWIN F. BECKENBACH

AND

RICHARD BELLMAN

WITH 6 FIGURES

SPRINGER-VERLAG
BERLIN · GÖTTINGEN · HEIDELBERG
1961

Printed in Germany

Druck der Brühlschen Universitätsdruckerei Gießen

TO

G. H. HARDY, J. E. LITTLEWOOD, AND G. PÓLYA

FROM TWO FOLLOWERS AFAR

Preface

Since the classic work on inequalities by HARDY, LITTLEWOOD, and PÓLYA in 1934, an enormous amount of effort has been devoted to the sharpening and extension of the classical inequalities, to the discovery of new types of inequalities, and to the application of inqualities in many parts of analysis. As examples, let us cite the fields of ordinary and partial differential equations, which are dominated by inequalities and variational principles involving functions and their derivatives; the many applications of linear inequalities to game theory and mathematical economics, which have triggered a renewed interest in convexity and moment-space theory; and the growing uses of digital computers, which have given impetus to a systematic study of error estimates involving much sophisticated matrix theory and operator theory.

The results presented in the following pages reflect to some extent these ramifications of inequalities into contiguous regions of analysis, but to a greater extent our concern is with inequalities in their native habitat. Since it is clearly impossible to give a connected account of the burst of analytic activity of the last twenty-five years centering about inequalities, we have decided to limit our attention to those topics that have particularly delighted and intrigued us, and to the study of which we have contributed.

We have tried to furnish a sufficient number of references to allow the reader to pursue a subject backward in time or forward in complexity, but we have made no attempt to be encyclopedic in covering a field either in the text or in the bibliography at the end of the separate chapters.

As with most authors, we have imposed upon our friends. To KY FAN we extend our sincere gratitude for reading the manuscript through several times and for furnishing us the most detailed suggestions. For the reading of individual chapters and for many valuable comments and references, we wish to thank R. P. BOAS, P. LAX, L. NIRENBERG, I. OLKIN, and O. TAUSSKY.

Our hope is that the reading of this book will furnish as much pleasure to others as the writing did to us.

Los Angeles and Santa Monica, 1961

EDWIN F. BECKENBACH
RICHARD BELLMAN

Contents

Chapter 1

The Fundamental Inequalities and Related Matters

§ 1. Introduction

In this initial chapter, we shall present many of the fundamental results and techniques of the theory of inequalities. Some of the results are important in themselves, and some are required for use in subsequent chapters; others are included, as are multiple proofs, on the basis of their elegance and unusual flavor [1].

We shall begin with the Cauchy inequality and the Lagrange identity, both of which will be substantially extended in this and the following chapter. From this we turn to a topic to which a monograph could be devoted in itself — namely, the famous inequality connecting the arithmetic and geometric means of n nonnegative numbers. Twelve proofs will be given of this basic result, not to suggest any lack of confidence in any single proof but rather to illustrate the wide range of techniques that the algebraist and analyst have at their disposal in treating inequalities. Of particular interest are the proofs of CAUCHY, HURWITZ, and BOHR.

Leaving this topic, albeit reluctantly, we shall establish the workhorses of analysis, the inequalities of HÖLDER and MINKOWSKI, in both discrete and continuous versions.

Subsequently, we shall establish some related, but more complex, results of BECKENBACH and DRESHER. These will be obtained with the aid of the important technique of quasi linearization, a method initiated by MINKOWSKI, developed by MAHLER, and used by YOUNG, ZYGMUND, and BELLMAN.

From this, we jump to the transformations of SCHUR involving doubly stochastic matrices, and to some results of KARAMATA, OSTROWSKI, and HARDY, LITTLEWOOD, and PÓLYA, pertaining to majorizing sequences. Continuous versions due to FAN and LORENTZ are also mentioned.

Our next port of call is in the domain of the elementary symmetric functions. Here, the results of MARCUS and LOPES are considerably more difficult to establish than might be suspected. Perhaps the most elegant proof of their inequalities is one that rests on the Minkowski theory of mixed volumes, a theory we shall discuss at length in our second volume on inequalities. Results due to WHITELEY are also presented.

From these matters, we turn to the fascinating questions of converses and refinements of the classical inequalities. Rather than follow the methods of BLASCHKE and PICK, and of BÜCKNER, or use moment-space arguments (the principal content of Chapter 3), we shall employ a method based on differential equations due to BELLMAN for establishing converse results. As far as the refinements are concerned, we shall merely mention some results and refer the reader to the original sources.

The last part of the chapter is devoted to some inequalities involving terms with alternating signs, discussed by WEINBERGER, SZEGÖ, OLKIN, BELLMAN, and others, all of which turn out to be particular cases of a novel inequality of STEFFENSEN.

§ 2. The Cauchy Inequality

The most basic inequality is the one stating that the square of any real number is nonnegative. To make effective use of this statement, we choose as our real number the quantity $y_1 - y_2$, where y_1 and y_2 are real. Then the inequality $(y_1 - y_2)^2 \geqq 0$ yields, upon multiplying out,

$$y_1^2 + y_2^2 \geqq 2 y_1 y_2. \tag{1}$$

The sign of equality holds if and only if $y_1 = y_2$. This is the simplest version of the inequality connecting the arithmetic and geometric means; following CAUCHY, we shall subsequently base one proof of the full result on this.

To make more effective use of the nonnegativity of squares, we form the sum

$$\sum_{i=1}^{n} (x_i u + y_i v)^2 = u^2 \sum_{i=1}^{n} x_i^2 + 2uv \sum_{i=1}^{n} x_i y_i + v^2 \sum_{i=1}^{n} y_i^2 , \tag{2}$$

where all quantities involved are real.

Since the foregoing quadratic form in u and v is nonnegative for all real values of u and v, its discriminant must be nonnegative, a fact expressed by the *Cauchy inequality* [1]:

$$\left(\sum_{i=1}^{n} x_i y_i\right)^2 \leqq \left(\sum_{i=1}^{n} x_i^2\right) \left(\sum_{i=1}^{n} y_i^2\right) . \tag{3}$$

This inequality may be considered as expressing the result that, in euclidean space of any number of dimensions, the cosine of an angle is less than or equal to 1 in absolute value. Equality holds if and only if the sets (x_i) and (y_i) are proportional, that is, if and only if there are numbers λ and μ, not both 0, such that

$$\lambda x_i + \mu y_i = 0 , \qquad i = 1, 2, \ldots, n.$$

Still more general results can be obtained by applying the foregoing argument not merely to an n-dimensional euclidean space, but to a general linear space S possessing an inner product for any two elements x and y, written (x, y), with the following properties:

$$\begin{aligned} &\text{(a)} \quad (x, x) \geqq 0 \text{ for each } x \in S, \\ &\text{(b)} \quad (x, y) = (y, x), \\ &\text{(c)} \quad (x, uy + vw) = u(x, y) + v(x, w) \text{ for all real scalars } u \text{ and } v. \end{aligned} \tag{4}$$

These properties enable us to conclude that the quadratic form in u and v,

$$(ux + vy, ux + vy) = u^2(x, x) + 2uv\,(x, y) + v^2(y, y)\,, \tag{5}$$

is nonnegative for all real u and v.

Hence, as above, we obtain the inequality

$$(x, y)^2 \leqq (x, x)\,(y, y)\,, \tag{6}$$

a result that is, in turn, a particular case of more general results we shall derive in Chapter 2; see § 2.6.

A large number of results may now be obtained in a routine way by a choice of S and the inner product (x, y). Thus, we may take

$$(x, y) = \int_a^b x(t)\, y(t)\, dG(t)\,, \tag{7}$$

a Riemann-Stieltjes integral with $G(t)$ nondecreasing for $a \leqq t \leqq b$, or

$$(x, y) = \sum_{i,j=1}^{n} a_{ij} x_i y_j, \tag{8}$$

where $A = (a_{ij})$ is a positive definite matrix, and so on.

§ 3. The Lagrange Identity

A problem of much interest and difficulty with surprising ramifications is that of replacing any given valid inequality by an identity that makes the inequality obvious. The inequality (2.3) can be derived immediately from the identity

$$\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right) - \left(\sum_{i=1}^{n} x_i y_i\right)^2 = \sum_{\substack{i,j=1 \\ i<j}}^{n} (x_i y_j - x_j y_i)^2. \tag{1}$$

This also is a special case of a more general identity discussed in § 6 of Chapter 2.

§ 4. The Arithmetic-mean — Geometric-mean Inequality

We shall begin our consideration of results less on the surface by discussing what is probably the most important inequality, and certainly a keystone of the theory of inequalities — namely, the arithmetic-mean — geometric-mean inequality. The result, of singular elegance, follows:

Theorem 1. *Let $x_1 x_2, \ldots, x_n$ be a set of n nonnegative quantities, $n \geqq 1$. Then*

$$\frac{x_1 + x_2 + \cdots + x_n}{n} \geqq (x_1 x_2 \ldots x_n)^{1/n}. \tag{1}$$

There is strict inequality unless the x_i are all equal.

Twelve proofs of this basic result will be presented in §§ 5—16, each based on a different principle or at least using a different device. There are a number of extensions of (1), involving weights. Amusingly enough, they are actually particularizations of the inequality, together with limiting cases. See § 14, below; a full discussion will be found also in [1.1].

§ 5. Induction — Forward and Backward

The following classical proof of Theorem 1 is due to Cauchy [2.1]. As noted in (2.1), for any two quantities y_1 and y_2 we have

$$y_1^2 + y_2^2 \geqq 2 y_1 y_2 . \tag{1}$$

Setting $y_1^2 = x_1$, $y_2^2 = x_2$ in this last inequality, we obtain

$$\frac{x_1 + x_2}{2} \geqq \sqrt{x_1 x_2} , \tag{2}$$

valid for any two nonnegative quantities x_1 and x_2. Referring to (2.1), we see that equality holds if and only if $x_1 = x_2$.

Now replace x_1 by the new variable $(x_1 + x_2)/2$, and x_2 by $(x_3 + x_4)/2$. Then (2), together with its repetition, yields

$$\begin{aligned} \frac{x_1 + x_2 + x_3 + x_4}{4} &\geqq \left[\frac{(x_1 + x_2)}{2} \frac{(x_3 + x_4)}{2}\right]^{1/2} \\ &\geqq \left[(x_1 x_2)^{1/2} (x_3 x_4)^{1/2}\right]^{1/2} = (x_1 x_2 x_3 x_4)^{1/4}. \end{aligned} \tag{3}$$

Proceeding in this way, we readily see that we can establish the inequality (4.1) for $n = 1, 2, 4, \ldots$, and, generally, for n a power of 2. This is a *forward* induction.

Let us now use *backward* induction. We shall show that if the inequality holds for n, then it holds for $n - 1$. In (4.1), replace x_n by the value

$$x_n = \frac{x_1 + x_2 + \cdots + x_{n-1}}{n - 1}, \tag{4}$$

$n \geqq 2$, and leave the other x_i unchanged. Then, from (4.1), we obtain the inequality

$$\begin{aligned} &\frac{x_1 + x_2 + \cdots + x_{n-1} + \dfrac{x_1 + x_2 + \cdots + x_{n-1}}{n - 1}}{n} \\ &\geqq (x_1 x_2 \ldots x_{n-1})^{1/n} \left(\frac{x_1 + x_2 + \cdots + x_{n-1}}{n - 1}\right)^{1/n}, \end{aligned} \tag{5}$$

or

$$\frac{x_1+x_2+\cdots+x_{n-1}}{n-1} \geqq (x_1 x_2 \ldots x_{n-1})^{1/n} \left(\frac{x_1+x_2+\cdots+x_{n-1}}{n-1}\right)^{1/n}. \tag{6}$$

Simplifying, we obtain

$$\left(\frac{x_1+x_2+\cdots+x_{n-1}}{n-1}\right) \geqq (x_1 x_2 \ldots x_{n-1})^{1/(n-1)}, \tag{7}$$

the desired inequality.

Combining the result for powers of 2 with this last result, we have an inductive proof of the theorem.

It is easy to see that the statement concerning strict inequality can also be established inductively.

Another interesting inequality that can be established by forward and backward induction is the following unpublished result due to KY FAN:

"If $0 < x_i \leqq 1/2$ for $i = 1, 2, \ldots, n$, then

$$\frac{\prod_{i=1}^{n} x_i}{\left(\sum_{i=1}^{n} x_i\right)^n} \leqq \frac{\prod_{i=1}^{n} (1-x_i)}{\left[\sum_{i=1}^{n} (1-x_i)\right]^n}, \tag{8}$$

with equality only if all the x_i are equal."

§ 6. Calculus and Lagrange Multipliers

Let us now approach the arithmetic-mean — geometric-mean inequality as a problem in calculus. We wish to minimize the function $x_1+x_2+\cdots+x_n$ over all nonnegative x_i satisfying the normalizing condition

$$x_1 x_2 \ldots x_n = 1. \tag{1}$$

Since the minimum clearly is not assumed at a boundary point, we can utilize the Lagrange-multiplier approach to determine the local minima. For the function

$$f(x_1, x_2, \ldots, x_n) = x_1 x_2 \ldots x_n - \lambda (x_1+x_2+\cdots+x_n), \tag{2}$$

the variational equations

$$\frac{\partial f}{\partial x_i} = \frac{x_1 x_2 x_3 \ldots x_n}{x_i} - \lambda = 0, \quad i = 1, 2, \ldots, n, \tag{3}$$

yield the result that $x_1 = x_2 = \cdots = x_n$. From this we readily see that $x_i = 1/n$, $i = 1, 2, \ldots, n$, is the unique minimizing point, and thus we obtain the inequality (4.1).

§ 7. Functional Equations

Theorem 1 can also be established through the functional-equation approach of dynamic programming [1]. We begin with the problem of maximizing $x_1 x_2 \ldots x_n$ subject to the constraints

$$x_1 + x_2 + \cdots + x_n = a, \ x_i \geqq 0 .$$

Denote this maximum value by $f_n(a)$, for $n = 1, 2, \ldots,$ and $a \geqq 0$. In order to obtain a recurrence relationship connecting the functions $f_n(a)$ and $f_{n-1}(a)$, we observe that once x_n has been chosen, the problem that remains is that of choosing $x_1, x_2, \ldots, x_{n-1}$ subject to the constraints

$$x_1 + x_2 + \cdots + x_{n-1} = a - x_n , \quad x_i \geqq 0 , \tag{1}$$

so as to maximize the product $x_1 x_2 \ldots x_{n-1}$.

It follows that

$$f_n(a) = \max_{0 \leqq x_n \leqq a} [x_n f_{n-1}(a - x_n)] , \quad n = 2, 3, \ldots, \tag{2}$$

with $f_1(a) = a$.

The change of variable $x_i = a y_i$, $i = 1, 2, \ldots, n$, enables us to conclude that

$$f_n(a) = a^n f_n(1) . \tag{3}$$

Using this functional form in (2), we see that

$$f_n(1) = f_{n-1}(1) \left[\max_{0 \leqq y \leqq 1} y (1 - y)^{n-1} \right] = \frac{f_{n-1}(1)\,(n-1)^{n-1}}{n^n} . \tag{4}$$

Since $f_1(1) = 1$, it follows that $f_n(1) = 1/n^n$, which is equivalent to (4.1).

§ 8. Concavity

Let us now present a proof of Theorem 1 by means of a geometric argument [1, 2, 3, 4]. Consider the curve $y = \log x$, shown in Fig. 1. Differentiation shows that the curve is concave, so that the chord joining any two of its points lies beneath the curve. Hence, for $x_1, x_2 > 0$,

$$\log \left(\frac{x_1 + x_2}{2} \right) \geqq \frac{\log x_1 + \log x_2}{2} , \tag{1}$$

with strict inequality unless $x_1 = x_2$.

This result is equivalent to

$$\frac{x_1 + x_2}{2} \geqq \sqrt{x_1 x_2} . \tag{2}$$

The same reasoning shows (see page 17) that

$$\log\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right) \geqq \frac{\log x_1 + \log x_2 + \cdots + \log x_n}{n}, \tag{3}$$

for $x_1, x_2, \ldots, x_n > 0$, and, generally, that

$$\log \frac{\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n}{\lambda_1 + \lambda_2 + \cdots + \lambda_n} \geqq \frac{\lambda_1 \log x_1 + \lambda_2 \log x_2 + \cdots + \lambda_n \log x_n}{\lambda_1 + \lambda_2 + \cdots + \lambda_n}, \tag{4}$$

for any combination of values $x_i \geqq 0$, $\lambda_i > 0$.

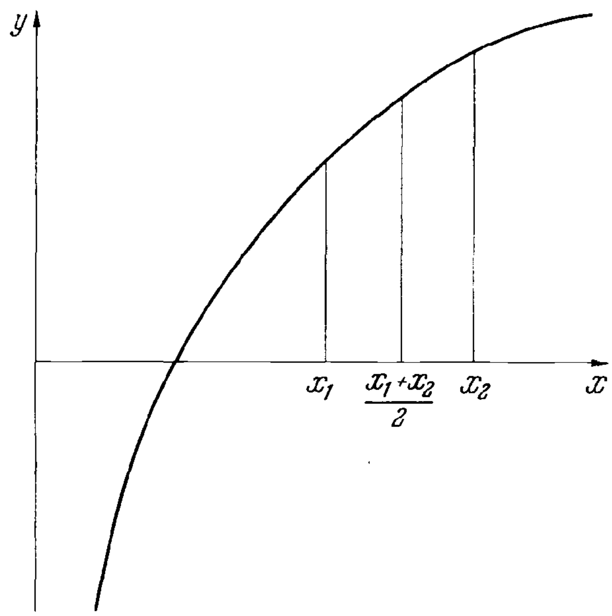

Fig. 1

This appears to be a stronger result than Theorem 1, but, as remarked in § 4, it can actually be obtained from (4.1) by specializing the values of the x_i and employing a limiting process; see §§ 14 and 16, below.

§ 9. Majorization — The Proof of Bohr

An amusing proof of Theorem 1 is due to H. Bohr [1].

To begin with, let us introduce the concept of majorization. Let $f(y)$ and $g(y)$ be two formal power series,

$$f(y) = \sum_{n=0}^{\infty} a_n y^n, \quad g(y) = \sum_{n=0}^{\infty} b_n y^n, \tag{1}$$

where $a_n, b_n \geqq 0$ for $n \geqq 0$.

If $a_n \geqq b_n$ for $n \geqq 0$, we write

$$f(y) \gg g(y)\,. \tag{2}$$

If $f_1(y) \gg g_1(y)$ and $f_2(y) \gg g_2(y)$, then clearly $f_1(y)\, f_2(y) \gg g_1(y)\, g_2(y)$.

Beginning with the obvious relationship

$$e^{xy} \gg \frac{x^N y^N}{N!}\,, \tag{3}$$

for $N = 1, 2, \ldots$, and $x, y \geqq 0$, we obtain

$$e^{y \sum_{i=1}^{n} x_i} \gg \frac{(x_1 x_2 \ldots x_n)^N y^{nN}}{(N!)^n}\,. \tag{4}$$

Hence, comparing the coefficients of y^{nN}, we get

$$\frac{\left(\sum_{i=1}^{n} x_i\right)^{nN}}{(nN)!} \geqq \frac{(x_1 x_2 \ldots x_n)^N}{(N!)^n}\,, \tag{5}$$

or

$$\frac{\left(\sum_{i=1}^{n} x_i\right)^{n}}{x_1 x_2 \ldots x_n} \geqq \left[\frac{(nN)!}{(N!)^n}\right]^{1/N} \tag{6}$$

for all positive integers N.

Since, as $k \to \infty$, we have Stirling's formula,

$$k! \sim k^k e^{-k} \sqrt{2\pi k}\,, \tag{7}$$

we see that

$$\lim_{N \to \infty} \left[\frac{(nN)!}{(N!)^n}\right]^{1/N} = n^n\,. \tag{8}$$

From (6) and (8) we obtain Theorem 1. This is the only proof we shall give that does not yield the condition under which the sign of equality holds.

§ 10. The Proof of Hurwitz

Let us now present an interesting proof due to Hurwitz [1]. This result was published in 1891, six years before his famous paper on the generation of invariants by integration over groups [2], and one may see the germ of the later technique in his earlier analysis, which follows.

For the function $f(x_1, x_2, \ldots, x_n)$ of the n real variables $x_1, x_2, \ldots, x_n$, let us denote by $Pf(x_1, x_2, \ldots, x_n)$ the sum of f over the $n!$ quantities that result from all possible $n!$ permutations of the x_i. Thus

$$\begin{aligned} P x_1^n &= (n-1)!\,(x_1^n + x_2^n + \cdots + x_n^n)\,, \\ P x_1 x_2 \ldots x_n &= n!\, x_1 x_2 \ldots x_n\,. \end{aligned} \tag{1}$$

Consider the functions ϕ_k, $k = 1, 2, \ldots, n-1$, obtained in the following manner:

$$\begin{aligned}
\phi_1 &= P\,[(x_1^{n-1} - x_2^{n-1})\,(x_1 - x_2)]\,,\\
\phi_2 &= P\,[(x_1^{n-2} - x_2^{n-2})\,(x_1 - x_2)\,x_3]\,,\\
\phi_3 &= P\,[(x_1^{n-3} - x_2^{n-3})\,(x_1 - x_2)\,x_3 x_4]\,,\\
&\;\vdots\\
\phi_{n-1} &= P\,[(x_1 - x_2)\,(x_1 - x_2)\,x_3 x_4 \ldots x_n]\,.
\end{aligned} \tag{2}$$

We see that

$$\begin{aligned}
\phi_1 &= P x_1^n + P x_2^n - P x_1^{n-1} x_2 - P x_2^{n-1} x_1\\
&= 2 P x_1^n - 2 P x_1^{n-1} x_2\,.
\end{aligned} \tag{3}$$

Similarly,

$$\begin{aligned}
\phi_2 &= 2 P x_1^{n-1} x_2 - 2 P x_1^{n-2} x_2 x_3\,,\\
\phi_3 &= 2 P x_1^{n-2} x_2 x_3 - 2 P x_1^{n-3} x_2 x_3 x_4\,,\\
&\;\vdots\\
\phi_{n-1} &= 2 P x_1^2 x_2 x_3 \ldots x_{n-1} - 2 P x_1 x_2 \ldots x_n\,.
\end{aligned} \tag{4}$$

Adding these results, we have

$$\phi_1 + \phi_2 + \cdots + \phi_{n-1} = 2 P x_1^n - 2 P x_1 x_2 \ldots x_n\,, \tag{5}$$

or, referring to (1),

$$\frac{x_1^n + x_2^n + \cdots + x_n^n}{n} - x_1 x_2 \ldots x_n = \frac{1}{2n!}\,(\phi_1 + \phi_2 + \cdots + \phi_n)\,. \tag{6}$$

It is easy to see that each of the functions $\phi_k(x)$ is nonnegative for $x_i \geqq 0$, since

$$\begin{aligned}
\phi_k &= P\,[(x_1^{n-k} - x_2^{n-k})\,(x_1 - x_2)\,x_3 x_4 \ldots x_{k+1}]\\
&= P\,[(x_1 - x_2)^2\,(x_1^{n-k-1} + \cdots + x_2^{n-k-1})\,x_3 x_4 \ldots x_{k+1}]\,.
\end{aligned} \tag{7}$$

Thus the difference appearing on the left-hand side of the identity (6) is nonnegative, whence Theorem 1 follows. This is the only proof we shall give that establishes the inequality (4.1) by means of an appropriate identity.

§ 11. A Proof of Ehlers

We shall prove Theorem 1 by showing that

$$x_1 x_2 \ldots x_n = 1,\; x_i \geqq 0,$$

implies that

$$x_1 + x_2 + \cdots + x_n \geqq n.$$

Assume that the result is valid for n, and let

$$x_1 x_1 \ldots x_n x_{n+1} = 1.$$

Let x_1 and x_2 be two of the x_i with the property that $x_1 \geqq 1$ and $x_2 \leqq 1$.

Then we have $(x_1 - 1)(x_2 - 1) \leqq 0$, or

$$x_1 x_2 + 1 \leqq x_1 + x_2 . \tag{1}$$

Hence

$$x_1 + x_2 + x_3 + \cdots + x_{n+1} \geqq 1 + x_1 x_2 + x_3 + \cdots + x_{n+1} \geqq 1 + n , \tag{2}$$

by the inequality for the n quantities $x_1 x_2, x_3, \ldots, x_n, x_{n+1}$. Since the result is trivial for $n = 1$, the validity of Theorem 1 follows. See [1, 2].

§ 12. The Arithmetic-Geometric Mean of Gauss; the Elementary Symmetric Functions

Let a_0, b_0 be two positive numbers with $a_0 \geqq b_0$, and define the further elements of the sequences $\{a_n\}$, $\{b_n\}$ as follows:

$$a_{n+1} = \frac{a_n + b_n}{2}, \quad b_{n+1} = (a_n b_n)^{1/2}. \tag{1}$$

It is easy to see that

$$a_0 \geqq a_1 \geqq \cdots \geqq a_n \geqq \cdots \geqq b_n \geqq \cdots \geqq b_1 \geqq b_0 , \tag{2}$$

and it can be shown that the sequences $\{a_n\}$ and $\{b_n\}$ have a common limit $M(a_0, b_0)$. This function $M(a_0, b_0)$ was first investigated by Gauss [1]. It plays an important role in the theory of elliptic functions, and, indeed, Gauss showed how the theory could be founded on this function.

The foregoing result concerning convergence of the sequences $\{a_n\}$ and $\{b_n\}$ can be greatly extended. For example, if $a_0 \geqq b_0 \geqq c_0 > 0$, and

$$a_{n+1} = \frac{a_n + b_n + c_n}{3}, \quad b_{n+1} = \left(\frac{a_n b_n + a_n c_n + b_n c_n}{3}\right)^{1/2}, \quad c_{n+1} = (a_n b_n c_n)^{1/3}, \tag{3}$$

it is easy to show that

$$\lim_{n \to \infty} a_n = \lim_{n \to \infty} b_n = \lim_{n \to \infty} c_n = M(a_0, b_0, c_0); \tag{4}$$

see Schapira [2], Schlesinger [3], and Bellman [4], where many further results concerning symmetric means are established.

One way to establish results concerning symmetric means is to use a set of interesting inequalities connecting the elementary symmetric functions of n real quantities. It turns out that the arithmetic-mean — geometric-mean inequality is merely one of a chain of inequalities.

Following the presentation in [1.1], we use a method based on Rolle's theorem. This method shows that valuable consequences can be derived from knowing that all the roots of a given polynomial equation are real, as well as from knowing, as in § 2, that there are no real roots. The same theme will be developed subsequently in the presentation of some results due to Garding; see §§ 36—38, below.

The result we require is the following immediate consequence of ROLLE's theorem:

Lemma. *If all the roots x/y of the equation*

$$f(x, y) \equiv c_0 x^m + c_1 x^{m-1} y + \cdots + c_m y^m = 0 \tag{1}$$

are real, then the same is true of all the equations of positive degree derived from it by partial differentiation with respect to x and y.

Let us apply this lemma to the polynomial

$$f(x, y) = (x + r_1 y)(x + r_2 y) \dots (x + r_n y), \tag{2}$$

where the r_i are real. Writing

$$f(x, y) = x^n + p_1 \binom{n}{1} x^{n-1} y + p_2 \binom{n}{2} x^{n-2} y^2 + \cdots + p_n y^n, \tag{3}$$

where

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}, \tag{4}$$

$p_0 = 1$, and (for $k = 1, 2, \dots, n-1$) p_k is the k-th elementary symmetric function, suitably weighted to give the average value of the products involved, we see that the equation

$$p_{k-1} x^2 + 2 p_k x y + p_{k+1} y^2 = 0, \tag{5}$$

obtained by repeated differentiation, has both its roots real. Thus we have

$$p_{k-1} p_{k+1} \leqq p_k^2, \tag{6}$$

$k = 1, 2, \dots, n-1$. Observe that this result holds for all real r_i, positive, negative, or zero.

For the next result, due to MACLAURIN [5], suppose that all the r_i are positive. Then from (6), we have

$$(p_0 p_2)(p_1 p_3)^2 (p_2 p_4)^3 \dots (p_{k-1} p_{k+1})^k \leqq p_1^2 p_2^4 \dots p_k^{2k}, \tag{7}$$

or

$$p_k^{1/k} \geqq p_{k+1}^{1/(k+1)}, \tag{8}$$

$k = 1, 2, \dots, n-1$.

From (8) we obtain

$$p_1 \geqq p_n^{1/n},$$

the arithmetic-mean — geometric-mean inequality.

§ 13. A Proof of Jacobsthal

There are a number of proofs of the arithmetic-mean — geometric-mean inequality based on algebraic relationships connecting these means. An interesting example is the following [1].

We begin with the identity

$$A_n = \frac{G_{n-1}}{n}\left[(n-1)\frac{A_{n-1}}{G_{n-1}} + \left(\frac{G_n}{G_{n-1}}\right)^n\right], \tag{1}$$

where

$$A_n = \frac{\sum_{i=1}^{n} x_i}{n}, \quad G_n = \left(\prod_{i=1}^{n} x_i\right)^{1/n}.$$

We next apply the inequality

$$z^n + n - 1 \geqq nz, \tag{2}$$

valid for $z \geqq 0$ and for $n \geqq 1$. For integral values of n, (2) follows easily from the identity

$$z^n - nz + n - 1 \equiv (z-1)(z^{n-1} + z^{n-2} + \cdots + z - n + 1).$$

If we take

$$z = \frac{G_n}{G_{n-1}},$$

then we obtain

$$A_n \geqq \frac{G_{n-1}}{n}\left[(n-1)\frac{A_{n-1}}{G_{n-1}} - (n-1) + \frac{nG_n}{G_{n-1}}\right], \tag{3}$$

or, simplifying,

$$A_n - G_n \geqq \frac{(n-1)}{n}(A_{n-1} - G_{n-1}). \tag{4}$$

The general result that $A_n - G_n \geqq 0$ follows inductively.

§ 14. A Fundamental Relationship

The remarkable inequalities [cf. (13.2)]

$$x^\alpha - \alpha x + \alpha - 1 \geqq 0, \quad \alpha > 1 \quad \text{or} \quad \alpha < 0, \tag{1}$$

$$x^\alpha - \alpha x + \alpha - 1 \leqq 0, \quad 0 < \alpha < 1, \tag{2}$$

which hold for $x > 0$, can be taken to be fundamental to the entire theory, for quite directly from them follow the arithmetic-mean — geometric-mean inequality and also the basic inequalities of Hölder and Minkowski; see §§ 17—18, below.

An easy application of differential calculus establishes (1) for $\alpha > 1$ or $\alpha < 0$, and (2) for $0 < \alpha < 1$; the sign of equality holds if and only if $x = 1$.

A longer but more elementary proof is the following [1.1]: For $y > 0$ and n a positive integer, the identity

$$\frac{y^{n+1} - 1}{n+1} - \frac{y^n - 1}{n} = \frac{y-1}{n(n+1)}(ny^n - y^{n-1} - \cdots - y - 1)$$

shows that

$$\frac{y^{n+1} - 1}{n+1} - \frac{y^n - 1}{n} \geqq 0,$$

the sign of equality holding if and only if $y = 1$. Hence, for any integer $m > n$,

$$\frac{y^m - 1}{m} - \frac{y^n - 1}{n} \geqq 0,$$

and therefore, with $y = x^{1/n}$ for any $x > 0$,

$$x^{m/n} - 1 - \frac{m}{n}(x-1) \geqq 0\,,$$

whence we obtain (1) for rational values $\alpha > 1$, namely

$$x^{m/n} - \frac{m}{n}x + \frac{m}{n} - 1 \geqq 0\,, \quad \frac{m}{n} > 1\,, \tag{3}$$

the sign of equality holding if and only if $x = 1$.

Now (1) follows from (3) for irrational $\alpha > 1$ as $m/n \to \alpha$, but in the limiting process the strict inequality is lost for $x \neq 1$. In order to regain it, let $\alpha = r\beta$, where r and β are both greater than 1 and r is rational. Then

$$x^\alpha - \alpha x + \alpha - 1 = (x^\beta)^r - r\beta x + r\beta - 1 > r x^\beta - r\beta x + r\beta - r \geqq 0\,,$$

and this completes the proof of (1) for $\alpha > 1$.

The substitution

$$x^\alpha = x^{1-\beta} = y^{\beta-1}, \quad \alpha > 1\,,$$

in (1) yields

$$y^{-1}(y^\beta - \beta y + \beta - 1) \geqq 0\,, \quad \beta < 0\,,$$

so that (1) holds also for $\alpha < 0$. Similarly, the substitution

$$x^\alpha = x^{1/\beta} = y\,, \quad \alpha > 1\,,$$

shows that (2) holds for $0 < \alpha < 1$. As before, the sign of equality holds in (1) for $\alpha < 0$ and in (2) for $0 < \alpha < 1$ if and only if $x = 1$. In the limiting cases $\alpha = 0$ and $\alpha = 1$, the sign of equality holds trivially for all $x > 0$:

$$x^\alpha - \alpha x + \alpha - 1 \equiv 0\,, \alpha = 0 \quad \text{or} \quad \alpha = 1\,.$$

To establish the arithmetic-mean — geometric-mean inequality, we note first that for positive x_1, x_2, the substitution

$$x = \frac{x_1}{x_2}$$

in (2) yields

$$\left(\frac{x_1}{x_2}\right)^\alpha - \alpha\frac{x_1}{x_2} + \alpha - 1 \leqq 0\,,$$

whence

$$x_1^\alpha x_2^{1-\alpha} \leqq \alpha x_1 + (1-\alpha)\,x_2\,, 0 < \alpha < 1\,,$$

so that the desired inequality holds for arbitrary x_1, $x_2 \geqq 0$ and arbitrary positive weights α, $1-\alpha$; the sign of equality holds if and only if $x_1 = x_2$.

Mathematical induction now readily gives the general result that

$$\prod_{i=1}^{n} x_i^{\alpha_i} \leqq \sum_{i=1}^{n} \alpha_i x_i \tag{4}$$

for

$$x_i \geqq 0\,, \alpha_i > 0\,, \sum_{i=1}^{n} \alpha_i = 1\,, \tag{5}$$

the sign of equality holding if and only if $x_i = x_j$ for all $i, j = 1, 2, \ldots, n$. Thus, if the result is assumed to hold for n, then for

$$x_i \geqq 0\,, \alpha_i > 0\,, \sum_{i=1}^{n+1} \alpha_i = 1\,,$$

we set

$$y_i = x_i\,, \beta_i = \alpha_i, i = 1, 2, \ldots, n-1\,,$$

and

$$y_n = x_n^{\alpha_n/\beta_n} x_{n+1}^{\alpha_{n+1}/\beta_n},\ \beta_n = \alpha_n + \alpha_{n+1}\,.$$

Now we have

$$y_i \geqq 0\,, \beta_i > 0\,, \sum_{i=1}^{n} \beta_i = 1\,,$$

and therefore, by the induction hypothesis, we obtain

$$\begin{aligned} \prod_{i=1}^{n+1} x_i^{\alpha_i} &= \prod_{i=1}^{n} y_i^{\beta_i} \\ &\leqq \sum_{i=1}^{n} \beta_i y_i \\ &= \sum_{i=1}^{n} \alpha_i x_i + (\alpha_n + \alpha_{n+1}) (x_n^{\alpha_n/\beta_n} x_{n+1}^{\alpha_{n+1}/\beta_n}) \\ &\leqq \sum_{i=1}^{n+1} \alpha_i x_i\,, \end{aligned}$$

the sign of equality holding throughout if and only if all the x_i are equal.

Thus we have again (cf. § 8) established the arithmetic-mean — geometric-mean inequality (4) for arbitrary x_i, α_i satisfying (5); but this time our proof for real (not necessarily rational) α_i has been about as elementary as possible.

The inequalities (1) and (2) are sometimes rewritten symmetrically by substituting a/b for x $(a > 0\,, b > 0)$ and letting

$$\alpha = \frac{1}{p}\,, 1 - \alpha = \frac{1}{q} \quad (p, q \neq 0 \text{ or } 1)\,.$$

Then

$$\begin{gathered} \frac{1}{p} + \frac{1}{q} = 1\,, q = \frac{p}{p-1}\,, p = \frac{q}{q-1}\,, p - 1 = \frac{p}{q}\,, \\ q - 1 = \frac{q}{p}\,, (p-1)(q-1) = 1\,; \end{gathered} \tag{6}$$

and (1) and (2) may be written equivalently as

$$a^{1/p}\, b^{1/q} \leqq \frac{a}{p} + \frac{b}{q} \text{ or } a^{1/p}\, b^{1/q} \geqq \frac{a}{p} + \frac{b}{q} \tag{7}$$

according as $p > 1$ or $p < 1$ $(p \neq 0)$. The sign of equality holds in the inequalities (7) if and only if $a = b$. It is easy to verify that for $p > 0$ the inequality (7) still holds under the slightly more general hypothesis that $a \geqq 0$, $b \geqq 0$.

§ 15. Young's Inequality

Let $y = \phi(x)$ be a continuous, strictly increasing function of x for $x \geqq 0$, with $\phi(0) = 0$. See Fig. 2. Examining the areas represented by the integrals, we see that

$$a b \leqq \int_0^a \phi(x)\, d x + \int_0^b \phi^{-1}(y)\, d y\,, \tag{1}$$

where $\phi^{-1}(y)$ is the function inverse to $\phi(x)$. It is easily seen that there is strict inequality unless $b = \phi(a)$. This is the inequality of YOUNG [1].

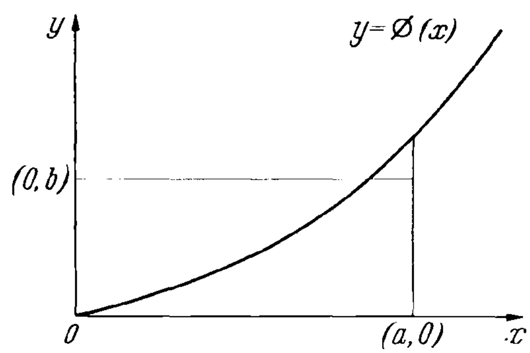

Fig. 2

Specializing ϕ, we obtain a number of interesting results.

With $y = x^{p-1}$, $p > 1$, (1) yields

$$a b \leqq \frac{a^p}{p} + \frac{b^q}{q}\,. \tag{2}$$

This is the first of the inequalities (14.7). From it we can readily obtain the other results given in § 14.

Choosing $y = \phi(x) = \log(x + 1)$ in YOUNG's inequality (1), and replacing a by $a - 1$, we obtain another interesting result, namely,

$$a b \leqq a \log a - a + e^b\,.$$

This inequality is frequently used in the theory of Fourier series.

§ 16. The Means $M_t(x, \alpha)$ and the Sums $S_t(x)$

In § 12 we saw that the arithmetic-mean — geometric-mean inequality is just one in a chain of inequalities involving the elementary symmetric functions. It is our purpose now to show how the arithmetic mean and the geometric mean fit also into a continuous hierarchy of mean values. Though elementary proofs could be given for the present results, we shall use differential calculus as our principal tool. We shall also use the theory of convex functions; in particular, we shall give an analytic justification of the geometric observations that were made in § 8. General discussions of convex functions and their applications have been given by BECKENBACH [8.3] and GREEN [8.4].

For any positive values

$$(x) \equiv (x_1, x_2, \ldots, x_n)$$

and positive weights

$$(\alpha) \equiv (\alpha_1, \alpha_2, \ldots, \alpha_n), \quad \sum_{i=1}^{n} \alpha_i = 1,$$

and any real $t \neq 0$, we define the *mean of order t*, or the *t norm*, of the values (x) with weights (α) by

$$M_t(x, \alpha) \equiv \left(\sum_{i=1}^{n} \alpha_i\, x_i^t \right)^{1/t}.$$

In particular, the means of order $-1, 1$, and 2 are the harmonic mean, the arithmetic mean, and the root-mean-square.

An easy application of l'Hospital's rule shows that

$$\lim_{t \to 0} M_t(x, \alpha) = \prod_{i=1}^{n} x_i^{\alpha_i}, \tag{1}$$

the geometric mean. Further, if $x_k = \max(x)$, then clearly

$$\alpha_k^{1/t} x_k \leqq M_t(x, \alpha) \leqq x_k$$

for $t > 0$, whence

$$\lim_{t \to \infty} M_t(x, \alpha) = \max(x). \tag{2}$$

Again, since

$$M_{-t}(x, \alpha) = \frac{1}{M_t(1/x, \alpha)},$$

we have

$$\lim_{t \to -\infty} M_t(x, \alpha) = \min(x). \tag{3}$$

Accordingly, we define

$$M_0(x, \alpha) = \prod_{i=1}^{n} x_i^{\alpha_i},$$

$$M_\infty(x, \alpha) = \max(x),$$

$$M_{-\infty}(x, \alpha) = \min(x).$$

If the x_i are assumed merely to be nonnegative, and at least one of the x_i is 0, then $M_t(x, \alpha)$ is taken to be 0 for $t \leqq 0$; but we shall consider only positive x_i.

We shall show that, *for positive values* x_i, $M_t(x, \alpha)$ *is a nondecreasing function of* t *for* $-\infty \leqq t \leqq \infty$, *and is strictly increasing unless all the* x_i *are equal.* This result includes the arithmetic-mean — geometric-mean inequality as a special case.

In order to verify the foregoing statement, we begin with an observation concerning convex functions. If a function $f(x)$ has a second derivative satisfying the inequality

$$\frac{d^2 f}{d x^2} > 0 \tag{4}$$

for $a < x < b$, then the graph $y = f(x)$ is convex in this interval. If the values

$$(x) \equiv (x_1, x_2, \ldots, x_n)$$

are all in the interval (a, b), and we let

$$\bar{x} = \sum_{i=1}^{n} \alpha_i x_i,$$

then $\bar{x}$ is also in (a, b), and by the mean-value theorem we have

$$f(x_i) = f(\bar{x}) + (x_i - \bar{x}) f'(\bar{x}) + \frac{(x_i - \bar{x})^2}{2} f''(\xi_i).$$

Multiplying by α_i and adding, we obtain

$$\sum_{i=1}^{n} \alpha_i f(x_i) = f(\bar{x}) + \sum_{i=1}^{n} \frac{\alpha_i (x_i - \bar{x})^2}{2} f''(\xi_i),$$

whence, by (4),

$$\sum_{i=1}^{n} \alpha_i f(x_i) \geqq f\left(\sum_{i=1}^{n} \alpha_i x_i\right), \tag{5}$$

the sign of equality holding if and only if all the x_i are equal; cf. § 8.

In particular, for the function

$$f(x) = x \log x, \quad x > 0,$$

we have

$$\frac{d^2 f}{d x^2} = \frac{1}{x} > 0,$$

so that by (5), for positive values (x),

$$\sum_{i=1}^{n} \alpha_i x_i \log x_i \geqq \left(\sum_{i=1}^{n} \alpha_i x_i\right) \log \sum_{j=1}^{n} \alpha_j x_j, \tag{6}$$

the sign of equality holding if and only if all the x_i are equal.

Now a computation yields

$$\frac{t^2}{M_t(x,\alpha)} \sum_{i=1}^{n} \alpha_i x_i^t \frac{dM_t(x,\alpha)}{dt} = \sum_{i=1}^{n} \alpha_i x_i^t \log x_i^t - \left(\sum_{i=1}^{n} \alpha_i x_i^t\right) \log \sum_{j=1}^{n} \alpha_j x_j^t ,$$

whence an application of (6) to the set of values (x^t) yields the desired result that

$$\frac{dM_t(x,\alpha)}{dt} \geqq 0 ,$$

the strict inequality holding unless all the x_i are equal.

Thus if not all the x_i are equal, the function M_t is a strictly increasing function of t and has two horizontal asymptotes. It might be expected, accordingly, that M_t has exactly one inflection value and thus that M_t is a convexo-concave function, but this is not necessarily the case [1]. Differentiation and an application of CAUCHY's inequality, however, shows that $t \log M_t(x, \alpha)$ is a convex function of t; accordingly, by (5), the function $M_t(x, \alpha)$ satisfies the inequality

$$M_T^T \leqq \prod_{i=1}^{n} M_{t_i}^{\alpha_i t_i} \tag{7}$$

for arbitrary t_i and for

$$T = \sum_{i=1}^{n} \alpha_i t_i , \ \alpha_i > 0 , \ \sum_{i=1}^{n} \alpha_i = 1 . \tag{8}$$

The *sum of order t*, defined by

$$S_t(x) = \left(\sum_{i=1}^{n} x_i^t\right)^{1/t} ,$$

behaves rather differently as a function of t. It decreases steadily from min (x) to 0 as t increases from $-\infty$ to $0-$, and decreases steadily from ∞ to max (x) as t increases from $0+$ to $+\infty$.

The inequality

$$S_{t_2}(x) \leqq S_{t_1}(x) , \ 0 < t_1 < t_2 ,$$

is sometimes called JENSEN's *inequality* [2,3], though this name is usually reserved for the inequality (5) — which holds for continuous convex functions generally, not just those having a positive second derivative.

From the fact that $t \log M_t(x, \alpha)$ is convex in t, it readily follows that $t \log S_t(x)$ also is a convex function of t, so that (7) still holds with S in place of M.

The function $S_t(x)$ is not necessarily concave for $t < 0$ [5]; but it is convex for $t > 0$ [4,5]. Accordingly, S_t satisfies the inequality

$$S_T \leqq \sum_{i=1}^{n} \alpha_i S_{t_i}$$

for arbitrary $t_i > 0$ and for T and α_i as in (8). In fact (a stronger result), $\log S_t$ is a convex function of t for $t > 0$, so that by (5) we have

$$S_T \leqq \prod_{i=1}^{n} S_{t_i}^{\alpha_i}.$$

§ 17. The Inequalities of Hölder and Minkowski

In (14.7) we saw that for

$$\frac{1}{p} + \frac{1}{q} = 1, \; p > 1,$$

we have

$$a^{1/p} b^{1/q} \leqq \frac{a}{p} + \frac{b}{q}, \tag{1}$$

and that the sign of inequality is reversed for $p < 1$ $(p \neq 0)$.

If we set successively

$$a = \frac{x_i^p}{X}, \; X = \sum_{i=1}^{n} x_i^p,$$

$$b = \frac{y_i^q}{Y}, \; Y = \sum_{i=1}^{n} y_i^q,$$

for $i = 1, 2, \ldots, n$, and add, we obtain the result

$$\frac{\sum_{i=1}^{n} x_i y_i}{X^{1/p} Y^{1/q}} \leqq \frac{1}{p} \frac{\sum_{i=1}^{n} x_i^p}{X} + \frac{1}{q} \frac{\sum_{i=1}^{n} y_i^q}{Y} = 1 \tag{2}$$

for $p > 1$, and the opposite inequality for $p < 1$ $(p \neq 0)$; equality holds if and only if the sets (x^p) and (y^q) are proportional. Thus we have established the classical inequality of Hölder [1]:

Theorem 2. *If $x_i, y_i \geqq 0$, $p > 1$, $1/p + 1/q = 1$, then*

$$\sum_{i=1}^{n} x_i y_i \leqq \left(\sum_{i=1}^{n} x_i^p\right)^{1/p} \left(\sum_{i=1}^{n} y_i^q\right)^{1/q}. \tag{3}$$

The inequality is reversed for $p < 1$ $(p \neq 0)$. (For $p < 0$, we assume that $x_i, y_i > 0$.) In each case, the sign of equality holds if and only if the sets (x^p) and (y^q) are proportional.

To complete the enumeration of the classical inequalities, let us add that due to Minkowski [2]:

Theorem 3. *If $x_i, y_i \geqq 0$, $p > 1$, then*

$$\left[\sum_{i=1}^{n} (x_i + y_i)^p\right]^{1/p} \leqq \left(\sum_{i=1}^{n} x_i^p\right)^{1/p} + \left(\sum_{i=1}^{n} y_i^p\right)^{1/p}. \tag{4}$$

The inequality is reversed for $p < 1$ ($p \neq 0$). (For $p < 0$, we assume that $x_i, y_i > 0$.) In each case, the sign of equality holds if and only if the sets (x) and (y) are proportional.

We shall first give a very short proof of this result, and then below, in § 20, show how to derive the result through quasi linearization.

Write

$$\sum_{i=1}^{n} (x_i + y_i)^p = \sum_{i=1}^{n} x_i (x_i + y_i)^{p-1} + \sum_{i=1}^{n} y_i (x_i + y_i)^{p-1}, \tag{5}$$

and apply HÖLDER's inequality with exponents p and q to each sum on the right. The result is

$$\sum_{i=1}^{n} (x_i + y_i)^p \leqq \left(\sum_{i=1}^{n} x_i^p\right)^{1/p} \left[\sum_{i=1}^{n} (x_i + y_i)^p\right]^{1/q} + \left(\sum_{i=1}^{n} y_i^p\right)^{1/p} \left[\sum_{i=1}^{n} (x_i + y_i)^p\right]^{1/q},$$

which is equivalent to (4). The sign of inequality is reversed for $p < 1$, $p \neq 0$. Equality holds if and only if the sets (x^p) and (y^p) are both proportional to $((x + y)^q)$, or equivalently if and only if the sets (x) and (y) are proportional to each other.

The inequality (4) is sometimes called the "triangle inequality" since, for $p = 2$, it is equivalent to the geometric inequality that in euclidean n space the sum of the lengths of two sides of a triangle is at least as great as the length of the third side. In this case $p = 2$, the inequality holds for all real, not necessarily positive, values of x_i, y_i, the condition for equality being that the sets (x) and (y) are positively proportional, that is, that there are numbers $\lambda \geqq 0$ and $\mu \geqq 0$, not both 0, such that

$$\lambda x_i = \mu y_i, \qquad i = 1, 2, \ldots, n.$$

§ 18. Extensions of the Classical Inequalities

The inequalities we have developed thus far admit many extensions and generalizations. In this section we shall briefly consider some of the more important of these.

Simple mathematical induction yields the following extensions of the inequalities of HÖLDER and MINKOWSKI, respectively:

If $x_{ij} \geqq 0$ for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$, and if $p_j > 1$ with $\sum_{j=1}^{n} 1/p_j = 1$, then

$$\sum_{i=1}^{n} \prod_{j=1}^{m} x_{ij} \leqq \prod_{j=1}^{m} \left(\sum_{i=1}^{n} x_{ij}^{p_j}\right)^{1/p_j}, \tag{1}$$

the sign of equality holding if and only if the m sets $(x_{i1}^{p_1}), (x_{i2}^{p_2}), \ldots, (x_{im}^{p_m})$ *are proportional, that is, if and only if there are numbers* λ_i, *not all* 0, *such that*

$$\sum_{j=i}^{m} \lambda_j \, x_{ij}^{p_j} = 0 \,,$$

for $i = 1, 2, \ldots, n$.

If $x_{ij} \geqq 0$ *for* $i = 1, 2, \ldots, n$ *and* $j = 1, 2, \ldots, m$, *and if* $p > 1$, *then*

$$\left[\sum_{i=1}^{n} \left(\sum_{j=1}^{m} x_{ij}\right)^{p}\right]^{1/p} \leqq \sum_{j=1}^{m} \left(\sum_{i=1}^{n} x_{ij}^{p}\right)^{1/p}. \tag{2}$$

The inequality is reversed for $p < 1$ $(p \neq 0)$. *(For* $p < 0$, *we assume that* $x_{ij} > 0$.) *In each case, the sign of equality holds if and only if the m sets* $(x_{i1}), (x_{i2}), \ldots (x_{im})$ *are proportional.*

Extensions to multiple sums and infinite sums can also be given. As pointed out in § 14, however, in any infinite process special care must be taken in treating conditions under which the sign of equality holds. Details can be found in [1.1].

Since the foregoing inequalities are "homogeneous in Σ," they admit mean-value analogues. Thus the analogue of (1) is

$$\frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{m} x_{ij} \leqq \prod_{j=1}^{m} \left(\frac{1}{n} \sum_{i=1}^{n} x_{ij}^{p_j}\right)^{1/p_j}, \quad p_j > 1, \sum \frac{1}{p_j} = 1 \,,$$

and factors $1/n$ or $1/m$ or both may be inserted at appropriate places in (2).

Quite generally, inequalities that are homogeneous in Σ also admit integral analogues. Thus the Hölder and Minkowski inequalities lead to the following result (the Cauchy inequality is the special case $p = 2$ of the discrete case of the Hölder inequality; its integral analogue is variously designated the Cauchy-Schwarz inequality, or the Schwarz inequality, or the Buniakowsky-Schwarz inequality):

Theorem 4. *Let* $f(P)$ *and* $g(P)$ *be functions defined for P in a region R, and let dV be a volume element in this region. Then, whenever the integrals on the right-hand sides of the inequalities exist, the integrals on the left-hand sides exist and satisfy the stated inequalities:*

$$\int_R fg \, dV \leqq \left(\int_R |f|^2 \, dV\right)^{1/2} \left(\int_R |g|^2 dV\right)^{1/2}, \tag{3}$$

(Buniakowsky-Schwarz);

$$\int_R fg \, dV \leqq \left(\int_R |f|^p \, dV\right)^{1/p} \left(\int_R |g|^q \, dV\right)^{1/q}, \quad p > 1 \,, \quad \frac{1}{p} + \frac{1}{q} = 1 \,, \tag{4}$$

(Hölder);

$$\left(\int_R |f + g|^p \, dV\right)^{1/p} \leqq \left(\int_R |f|^p \, dV\right)^{1/p} + \left(\int_R |g|^p \, dV\right)^{1/p}, \quad p > 1 \,, \tag{5}$$

(Minkowski).

The signs of equality hold in (3), (4), *and* (5) *if and only if the functions f and g are positively proportional (except at most on a set of measure zero).*

The Minkowski inequality can be further extended by replacing the sums in (5) by integrals:

$$\left[\int_R \left|\int_S f\,dV_S\right|^p dV_R\right]^{1/p} \leqq \int_S \left[\left(\int_R |f|^p\,dV_R\right)^{1/p}\right] dV_S\,,\ p > 1\,. \tag{6}$$

Or in (2) the sums with respect to j but not with respect to i might be replaced by integrals. In each case, the inequality is reversed for $p < 1$ ($p \neq 0$), but for $p < 0$ we assume that the functions are nowhere equal to zero.

There are several ways of demonstrating integral results of this nature. Either we can derive them as limiting forms of the discrete versions, or we can establish them directly, utilizing techniques that are equally applicable to discrete or continuous versions.

To illustrate the first technique, let us sketch a proof of the inequality

$$\left(\int_0^1 |fg|\,dx\right)^2 \leqq \left(\int_0^1 |f|^2\,dx\right)\left(\int_0^1 |g|^2\,dx\right). \tag{7}$$

To begin with, assume that f and g are continuous over $[0,1]$. Then (7) follows as the limiting form of

$$\left[\frac{1}{N}\sum_{k=0}^{N-1} f(k\Delta)\,g\,(k\Delta)\,\Delta\right]^2 \leqq \left[\frac{1}{N}\sum_{k=0}^{N-1} |f(k\Delta)|^2\,\Delta\right]\left[\frac{1}{N}\sum_{k=0}^{N-1} |g(k\Delta)|^2\,\Delta\right], \tag{8}$$

a consequence of the discrete inequality derived in § 2.

To obtain (7) in full generality, we use the fact that Lebesgue integrable functions can be approximated in L^1-norm by means of polynomials. It is clear that this mode of proof is not very elegant, and can lead to difficulties when the region R is quite general.

Let us then present an illustrative direct proof. For the Buniakowsky-Schwarz inequality, we can proceed as follows. For any two real values u and v we have

$$|u|^2 + |v|^2 \geqq 2\,|uv|\,. \tag{9}$$

Regarding u and v as functions of P and integrating over R, we have

$$\int_R |u|^2\,dV + \int_R |v|^2\,dV \geqq 2\int_R |uv|\,dV\,. \tag{10}$$

Replace u by $|f|\Big/\left(\int_R |f|^2\,dV\right)^{1/2}$ and v by $|g|\Big/\left(\int_R |g|^2\,dV\right)^{1/2}$. Then (10) yields

$$\frac{\int_R |f|^2 dV}{\int_R |f|^2 dV} + \frac{\int_R |g|^2 dV}{\int_R |g|^2 dV} \geqq 2\,\frac{\int_R |fg|\,dV}{\left(\int_R |f|^2\,dV\right)^{1/2}\left(\int_R |g|^2\,dV\right)^{1/2}}, \tag{11}$$

which implies (7).

Extensive generalizations of the Buniakowsky-Schwarz inequality may be found in the book by Bôcher [1] and the paper by Ogura [2]. The geometric interpretations given by Ogura for the function-space inequalities are analogous to those given by Bôcher for the Euclidean results.

We have pointed out that since the Hölder and Minkowski inequalities are homogeneous in Σ, they admit mean-value analogues and also admit integral analogues. For the same reason, they admit mean-value integral analogues. We have only to replace

$$\int_R \text{ by } \frac{1}{\operatorname{meas} R}\int_R$$

throughout. The inequality

$$S_r(x) \geqq S_t(x)\,, \quad r < t < 0 \quad \text{or} \quad 0 < r < t\,,$$

of § 14 concerning finite sums is not homogeneous in Σ and does not have an integral analogue. But the opposite inequality

$$M_r(x) \leqq M_t(x)\,, \quad -\infty \leqq r < t \leqq \infty\,,$$

concerning means, while also not homogeneous in Σ, does have an integral analogue:

$$\left(\frac{1}{\operatorname{meas} R}\int_R |f|^r dV\right)^{1/r} \leqq \left(\frac{1}{\operatorname{meas} R}\int_R |f|^t dV\right)^{1/t}.$$

Here, for $M_{-\infty}$ and M_∞ we understand the essential minimum and essential maximum (the infinum and supremum disregarding sets of measure zero), and for M_0 we have the limiting value, or geometric mean,

$$\exp\left(\frac{1}{\operatorname{meas} R}\int_R \log |f|\, dV\right).$$

§ 19. Quasi Linearization

Let us begin our discussion of quasi linearization with the observation that an equivalent statement of the discrete version of Hölder's inequality for $p > 1$ is the following:

Theorem 5. *For $x_i \geqq 0$, $p > 1$, we have the representation*

$$\left(\sum_{i=1}^{n} x_i^p\right)^{1/p} = \max_{R(y)} \sum_{i=1}^{n} x_i y_i\,, \tag{1}$$

where $R(y)$ is the region defined by

$$\sum_{i=1}^{n} y_i^q = 1\,, y_i \geqq 0\,. \tag{2}$$

The importance of this representation lies in the fact that we can represent a *nonlinear* function such as that appearing on the left side of (1) as an envelope of *linear* functions. Thus, we can establish a number of *nontrivial* properties of the nonlinear function as simple consequences of *trivial* properties of the linear function $\sum_{i=1}^{n} x_i y_i$. A more detailed discussion of this technique will be given below in §§ 25 and 26.

Let us now discuss some simple aspects of the above concept.

First, let $L(x, y)$ be a function of two variables x and y, where x and y are elements of normed spaces R and S, respectively. Let $\|y\|$ denote the norm of the element $y \in S$, and define the new function $\phi(x)$ of x alone by means of the relationship

$$\phi(x) = \max_{\|y\| \leqq 1} L(x, y) . \tag{3}$$

Simple functional properties of $L(x, y)$ as a function of x, valid for all y, such as positivity, linearity, and convexity, will be mirrored in corresponding properties of $\phi(x)$. In many cases, these properties are more readily demonstrated for $L(x, y)$ than are the corresponding properties for $\phi(x)$.

The first and most important case is that in which $L(x, y)$ is linear in x for all y, that is,

$$L(\alpha x_1 + \beta x_2, y) = \alpha L(x_1, y) + \beta L(x_2, y) . \tag{4}$$

From this, it follows that

$$\begin{aligned} \phi(x_1 + x_2) &= \max_{\|y\| \leqq 1} L(x_1 + x_2, y) \\ &= \max_{\|y\| \leqq 1} [L(x_1, y) + L(x_2, y)] \\ &\leqq \max_{\|y\| \leqq 1} L(x_1, y) + \max_{\|y\| \leqq 1} L(x_2, y) \\ &= \phi(x_1) + \phi(x_2) . \end{aligned} \tag{5}$$

This is the "triangle inequality", or the "subadditive property", for $\phi(x)$.

A second case of importance, which plays a central role in the succeeding chapter devoted to matrices (see also § 35 of the present chapter), is that in which

$$L(x, y) = \int_R e^{-M(x, y, z)} \, dG(y, z) , \tag{6}$$

where $dG \geqq 0$, the integration is over a region of z space, and $M(x, y, z)$ is linear in x for all y and z.

We have then, for $0 < \lambda < 1$,

$$L(\lambda x_1 + (1 - \lambda) x_2, y) = \int_R e^{-\lambda M(x_1, y, z)} e^{-(1-\lambda) M(x_2, y, z)} \, dG(y, z) . \tag{7}$$

Applying Hölder's inequality with exponents $p = 1/\lambda$, $q = 1/(1 - \lambda)$, we obtain the relationship

$$\begin{aligned} &L(\lambda x_1 + (1-\lambda)\, x_2, y) \\ &\leqq \Big(\int_R e^{-M(x_1,y,z)}\, dG\Big)^{\lambda} \Big(\int_R e^{-M(x_2,y,z)}\, dG\Big)^{1-\lambda} \\ &\leqq L(x_1, y)^{\lambda} L(x_2, y)^{1-\lambda}. \end{aligned} \tag{8}$$

Taking logarithms, we see that $\log L(x, y)$ is a convex function of x for all y.

In § 33 of Chapter 4, we shall employ a quasi linearization of the maximum functional; and in the book by BELLMAN, GLICKSBERG, and GROSS [1] there will be found some applications of the formula

$$|x| = \max_{|y| \leqq 1} xy \tag{9}$$

to the solution of unconventional problems in the calculus of variations. This technique was extensively used by ZYGMUND [2] in the theory of Fourier series.

§ 20. Minkowski's Inequality

As a first illustration of the foregoing type of argumentation, let us give another proof of the inequality of MINKOWSKI, established above in § 17.

Theorem 6. *For* $x_i, y_i \geqq 0, p > 1$, *we have*

$$\left[\sum_{i=1}^{n} (x_i + y_i)^p\right]^{1/p} \leqq \left(\sum_{i=1}^{n} x_i^p\right)^{1/p} + \left(\sum_{i=1}^{n} y_i^p\right)^{1/p}. \tag{1}$$

Proof. Since by HÖLDER's inequality we have

$$\left[\sum_{i=1}^{n} (x_i + y_i)^p\right]^{1/p} = \max_{R(z)} \sum_{i=1}^{n} (x_i + y_i)\, z_i\,, \tag{2}$$

where $R(z)$ is the region defined by

$$\sum_{i=1}^{n} z_i^q = 1\,,$$

it follows that

$$\begin{aligned} \left[\sum_{i=1}^{n} (x_i + y_i)^p\right]^{1/p} &\leqq \max_{R(z)} \sum_{i=1}^{n} x_i z_i + \max_{R(z)} \sum_{i=1}^{n} y_i z_i \\ &\leqq \left(\sum_{i=1}^{n} x_i^p\right)^{1/p} + \left(\sum_{i=1}^{n} y_i^p\right)^{1/p}, \end{aligned}$$

which is the stated result.

It is not difficult to show, by means of the condition for equality in HÖLDER's inequality, that there is equality in (1) if and only if either $p = 1$ or the sets (x) and (y) are proportional.

§ 21. Another Inequality of Minkowski

Using the same quasi-linearization techniques, let us establish a result we shall use below in §§ 34 and 35.

Theorem 7. *For* $x_i, y_i \geqq 0$, *we have*

$$\left[\prod_{i=1}^{n} (x_i + y_i)\right]^{1/n} \geqq \left(\prod_{i=1}^{n} x_i\right)^{1/n} + \left(\prod_{i=1}^{n} y_i\right)^{1/n}. \tag{1}$$

Proof. The arithmetic-mean — geometric-mean inequality asserts that

$$\left(\prod_{i=1}^{n} x_i\right)^{1/n} = \min_{R(z)} \sum_{i=1}^{n} \frac{x_i z_i}{n}, \tag{2}$$

where $R(z)$ is now the region defined by

$$\prod_{i=1}^{n} z_i = 1, \quad z_i \geqq 0. \tag{3}$$

Using (2), we have the desired result that

$$\begin{aligned} \left[\prod_{i=1}^{n} (x_i + y_i)\right]^{1/n} &= \min_{R(z)} \sum_{i=1}^{n} \frac{z_i (x_i + y_i)}{n} \\ &\geqq \min_{R(z)} \sum_{i=1}^{n} \frac{z_i x_i}{n} + \min_{R(z)} \sum_{i=1}^{n} \frac{z_i y_i}{n} \qquad (4) \\ &\geqq \left(\prod_{i=1}^{n} x_i\right)^{1/n} + \left(\prod_{i=1}^{n} y_i\right)^{1/n}. \end{aligned}$$

§ 22. Minkowski's Inequality for $0 < p < 1$

Again using quasi linearization, let us establish the following result:

Theorem 8. *If* $x_i, y_i \geqq 0$, $0 < p < 1$, *we have*

$$\left[\sum_{i=1}^{n} (x_i + y_i)^p\right]^{1/p} \geqq \left(\sum_{i=1}^{n} x_i^p\right)^{1/p} + \left(\sum_{i=1}^{n} y_i^p\right)^{1/p}. \tag{1}$$

Proof. Let $x_i = u_i^{1/p}$, $y_i = v_i^{1/p}$. Then we wish to prove that

$$\sum_{i=1}^{n} (u_i^{1/p} + v_i^{1/p})^p \geqq \left[\left(\sum_{i=1}^{n} u_i\right)^{1/p} + \left(\sum_{i=1}^{n} v_i\right)^{1/p}\right]^p. \tag{2}$$

Since for $0 < p < 1$, we have

$$\left[\left(\sum_{i=1}^{n} u_i\right)^{1/p} + \left(\sum_{i=1}^{n} v_i\right)^{1/p}\right]^p = \max_{R(z)} \left[z_1 \left(\sum_{i=1}^{n} u_i\right) + z_2 \left(\sum_{i=1}^{n} v_i\right)\right], \tag{3}$$

where $R(z)$ is defined by $z_1^q + z_2^q = 1$, $z_1, z_2 \geqq 0$, $q = 1/(1-p)$, it follows that

$$\left[\left(\sum_{i=1}^{n} u_i\right)^{1/p} + \left(\sum_{i=1}^{n} v_i\right)^{1/p}\right]^p = \max_{R(z)} \left[\sum_{i=1}^{n} (z_1 u_i + z_2 v_i)\right]$$

$$\leqq \sum_{i=1}^{n} \max_{R(z)} (z_1 u_i + z_2 u_i)$$

$$\leqq \sum_{i=1}^{n} (u_i^{1/p} + v_i^{1/p})^p ,$$

which is the desired result.

§ 23. An Inequality of Beckenbach

Let us now demonstrate [1] the following result:

Theorem 9. *Let* $1 \leqq p \leqq 2$, *and* $x_i, y_i > 0$ *for* $i = 1, 2, \ldots, n$; *then*

$$\frac{\sum_{i=1}^{n} (x_i + y_i)^p}{\sum_{i=1}^{n} (x_i + y_i)^{p-1}} \leqq \frac{\sum_{i=1}^{n} x_i^p}{\sum_{i=1}^{n} x_i^{p-1}} + \frac{\sum_{i=1}^{n} y_i^p}{\sum_{i=1}^{n} y_i^{p-1}} . \tag{1}$$

Proof. The inequality is trivial for $p = 1$. Otherwise, as in § 19, we have the representation

$$\left(\sum_{i=1}^{n} x_i^p\right)^{1/p} = \max_{R(z)} \sum_{i=1}^{n} x_i z_i , \quad p > 1 . \tag{2}$$

Consequently, it is sufficient to prove that

$$\frac{\left[\sum_{i=1}^{n} (x_i + y_i) z_i\right]^p}{\sum_{i=1}^{n} (x_i + y_i)^{p-1}} \leqq \frac{\left(\sum_{i=1}^{n} x_i z_i\right)^p}{\sum_{i=1}^{n} x_i^{p-1}} + \frac{\left(\sum_{i=1}^{n} y_i z_i\right)^p}{\sum_{i=1}^{n} y_i^{p-1}} , \tag{3}$$

for all $z_i \geqq 0$.

In order to simplify these expressions, let us set

$$x^p = \frac{\left(\sum_{i=1}^{n} x_i z_i\right)^p}{\sum_{i=1}^{n} x_i^{p-1}} , \quad y^p = \frac{\left(\sum_{i=1}^{n} y_i z_i\right)^p}{\sum_{i=1}^{n} y_i^{p-1}} . \tag{4}$$

Solving for the quantities $\sum_{i=1}^{n} x_i z_i$ and $\sum_{i=1}^{n} y_i z_i$, we can write (3) as

$$\frac{\left[x\left(\sum_{i=1}^{n} x_i^{p-1}\right)^{1/p} + y\left(\sum_{i=1}^{n} y_i^{p-1}\right)^{1/p}\right]^p}{\sum_{i=1}^{n} (x_i + y_i)^{p-1}} \leqq x^p + y^p . \tag{5}$$

To demonstrate the validity of inequality (5), we apply HÖLDER's inequality for $p > 1$ to the numerator of the left side, obtaining

$$\left[x\left(\sum_{i=1}^{n} x_i^{p-1}\right)^{1/p} + y\left(\sum_{i=1}^{n} y_i^{p-1}\right)^{1/p}\right]^{p} \tag{6}$$
$$\leq (x^p + y^p)\left[\left(\sum_{i=1}^{n} x_i^{p-1}\right)^{1/(p-1)} + \left(\sum_{i=1}^{n} y_i^{p-1}\right)^{1/(p-1)}\right]^{p-1}.$$

Hence, (5) holds if we have

$$\left(\sum_{i=1}^{n} x_i^{p-1}\right)^{1/(p-1)} + \left(\sum_{i=1}^{n} y_i^{p-1}\right)^{1/(p-1)} \tag{7}$$
$$\leq \left[\sum_{i=1}^{n} (x_i + y_i)^{p-1}\right]^{1/(p-1)}.$$

This, however, is MINKOWSKI's inequality, valid for $1 < p \leq 2$.

The inequality sign in (1) is reversed for $0 \leq p \leq 1$.

§ 24. An Inequality of Dresher

An extension of BECKENBACH's inequality was obtained by DRESHER [1] by means of moment-space techniques:

Theorem 10. *If $p \geq 1 \geq r \geq 0$, $f, g \geq 0$, then*

$$\left(\frac{\int |f+g|^p\, d\phi}{\int |f+g|^r\, d\phi}\right)^{1/(p-r)} \leq \left(\frac{\int f^p\, d\phi}{\int f^r\, d\phi}\right)^{1/(p-r)} \tag{1}$$
$$+ \left(\frac{\int g^p\, d\phi}{\int g^r\, d\phi}\right)^{1/(p-r)}.$$

This result can be derived through quasi linearization, as in § 23. It was also established by DANSKIN [2], who employed a combination of the Hölder and Minkowski inequalities.

§ 25. Minkowski-Mahler Inequality

The technique we have been using in the preceding §§ 19—24 is based on an idea introduced by MINKOWSKI [1]. Let $F(x) = F(x_1, x_2, \ldots, x_n)$ be a function possessing the following properties:

$$\begin{array}{lll} \text{(a)} & F(x) > 0\,, & \text{for } x \neq 0\,, \\ \text{(b)} & F(tx) = tF(x)\,, & \text{for } t \geq 0\,, \\ \text{(c)} & F(x) + F(y) \geq F(x+y)\,. & \end{array} \tag{1}$$

Clearly, $F(x)$ is a generalized distance function, or *norm*, associated with the n-dimensional vector x; see § 2.

Given a function satisfying the foregoing requirements, we can introduce a new function $G(y)$, the *polar function*, defined by the relationship

$$G(y) = \max_{x} \frac{(x, y)}{F(x)}\,. \tag{2}$$

In the theory of convex bodies (see BONNESEN and FENCHEL [8.1]), $G(y)$ is called the "Stützfunktion." It is defined geometrically by means of the transformation of reciprocal polars with respect to the sphere $(x, x) = 1$.

From this fact, we suspect that (2) is a reciprocal relationship in the sense that

$$F(x) = \max_y \frac{(x, y)}{G(y)} . \tag{3}$$

This was proved by MINKOWSKI; see page 24 of [8.1].

It follows that we have

$$(x, y) \leqq F(x)\, G(y) , \tag{4}$$

an inequality apparently first explicitly stated and used by MAHLER [2] in the geometrical theory of numbers.

Some interesting extensions of the pole-polar relationship are discussed in L. C. YOUNG [3], which we have followed in the preceding treatment; see also the paper by FENCHEL [4], which contains other references.

A detailed discussion of these matters is given in LORCH [5], [6], [7], [8.2], where many interesting results may be found, showing the intimate relation between inequalities, convexity, and the "mixed volume" concept of MINKOWSKI. These matters will be discussed in our second volume.

§ 26. Quasi Linearization of Convex and Concave Functions

The representation of $F(x)$ in the form contained in Equation (25.3) yields a quasi linearization that was used by L. C. YOUNG [25.3] in a fashion similar to our use of quasi linearization in §§ 19—24.

Let us now consider a quasi linearization that can be used in treating convex or concave functions and functionals that are not necessarily homogeneous. To begin with, let us treat the one-dimensional case. Take $f(u)$ to be a strictly convex function of u for all u, in the sense that $f''(u) > 0$. Then it is easy to see that

$$f(u) = \max_v \, [f(v) + (u - v) f'(v)] , \tag{1}$$

and that the unique maximum occurs at $v = u$. Similarly, for a strictly concave function we have

$$f(u) = \min_v \, [f(v) + (u - v) f'(v)] . \tag{2}$$

The general result is

Theorem 11. *Let $f(x) = f(x_1, x_2, \ldots, x_n)$ be a strictly convex function of x for all x; then*

$$f(x) = \max_y \, [f(y) + (x - y, \phi(y))] , \tag{3}$$

where $\phi(y) = (\partial f/\partial y_1, \partial f/\partial y_2, \ldots, \partial f/\partial y_n)$, *the gradient of* $f(y)$. *The unique maximum occurs at* $y = x$.

This type of quasi linearization has been extensively used by BELLMAN [1], [2] and KALABA [3] in connection with the analytic and computational treatment of nonlinear functional equations.

§ 27. Another Type of Quasi Linearization

Another kind of quasi linearization, dependent on the use of random variables, has been employed by ZYGMUND [1] and by ZYGMUND and MARCIENKIEWICZ [2].

In [2], they establish the following result:

Theorem 12. *Let* $S = \{f_i\}$ *be a linear family of functions belonging to* $L^p(a, b)$, $p > 0$, *and let* T *be a linear transformation with the property that* $Tf_i \in L^p(a, b)$ *for each* i. *If a constant* m *exists such that*

$$\int_a^b |Tf_i|^p \, dx \leqq m \int_a^b |f_i|^p \, dx \tag{1}$$

for all i, *then*

$$\int_a^b [\Sigma (Tf_i)^2]^{p/2} \, dx \leqq m \int (\Sigma f_i^2)^{p/2} \, dx \tag{2}$$

for any set of f_i *in* S.

A corresponding result for variations is given in [1].

§ 28. An Inequality of Karamata

As an application of the representation in Theorem 11, we shall establish a result due to OSTROWSKI [1], a generalization of the following one of KARAMATA [2]:

Theorem 13. *Suppose that we have* $2n$ *numbers*, $\{x_k, y_k\}$, $k = 1, 2, \ldots, n$, *satisfying the relations*

(a) $\quad x_1 \geqq x_2 \geqq \cdots \geqq x_n, \; y_1 \geqq y_2 \geqq \cdots \geqq y_n,$

(b)
$$\begin{aligned} &x_1 \geqq y_1, \\ &x_1 + x_2 \geqq y_1 + y_2, \\ &\vdots \\ &x_1 + x_2 + \cdots + x_n = y_1 = y_2 + \cdots y_n. \end{aligned} \tag{1}$$

Then for any continuous, convex function $\phi(x)$ *we have*

$$\phi(x_1) + \phi(x_2) + \cdots + \phi(x_n) \geqq \phi(y_1) + \phi(y_2) + \cdots + \phi(y_n). \tag{2}$$

Artificial as these relations may seem, it is nonetheless true that inequalities of this nature arise in an amazing variety of circumstances [3].

A proof of Theorem 13 is given in § 30.

§ 29. The Schur Transformation

KARAMATA obtained the result stated in the preceding section by use of the following theorem, of interest in itself.

Theorem 14. *A necessary and sufficient condition that* $2n$ *numbers* $\{x_k, y_k\}$ *be related by means of the inequalities in* (28.1) *is that*

$$y_k = \sum_{l=1}^{n} a_{kl} x_l, \quad k = 1, 2, \ldots, n, \tag{1}$$

where

$$\begin{aligned} &\text{(a)} \qquad a_{kl} \geqq 0 \,, \\ &\text{(b)} \qquad \sum_{l=1}^{n} a_{kl} = 1 \,, \\ &\text{(b)} \qquad \sum_{k=1}^{n} a_{kl} = 1 \,. \end{aligned} \tag{2}$$

Since transformations of this type were first discussed by SCHUR [1], we shall follow OSTROWSKI [28.1] in calling them *Schur transformations.* In more recent years, matrices (a_{kl}) of this type have been called *doubly — stochastic*. They play an important role in certain combinatorial problems; see BIRKHOFF [2]. A proof of this result may be found in HARDY, LITTLEWOOD, and PÓLYA [1.1], and a simpler proof in OSTROWSKI [28.1]; see also SCHREIBER [3], MIRSKY [4], RYSER [5], SCHUR [1], and FUCHS [31.1].

§ 30. Proof of the Karamata Result

Let us now indicate how quasi linearization may be used to establish the Karamata result of § 28. Our starting point is the representation

$$\phi(x_1) + \phi(x_2) + \cdots + \phi(x_n) = \max_{z} \left[\sum_{i=1}^{n} \phi(z_i) + \sum_{i=1}^{n} (x_i - z_i)\, \phi'(z_i) \right], \tag{1}$$

where we assume initially that $\phi(z)$ is strictly convex.

Since the maximum is assumed for $z_i = x_i$, and we suppose that $x_1 \leqq x_2 \leqq \cdots \leqq x_n$, it is sufficient to allow the z_i to vary only over the subset of all z defined by

$$z_1 \leqq z_2 \leqq \cdots \leqq z_n \,. \tag{2}$$

To establish the required inequality, we must show that for all $\{x_k, y_k\}$ satisfying (28.1), and all z_i satisfying (2), we have

$$\begin{aligned} &x_1\phi'(z_1) + x_2\phi'(z_2) + \cdots + x_n\phi'(z_n) \\ &\leqq y_1\phi'(z_1) + y_2\phi'(z_2) + \cdots + y_n\phi'(z_n) \,. \end{aligned} \tag{3}$$

This, however, is immediate from summation by parts, the Abel-Brunacci formula,

$$\begin{aligned} &x_1\phi'(z_1) + x_2\phi'(z_2) + \cdots + x_n\phi'(z_n) \\ &= s_1 [\phi'(z_1) - \phi'(z_2)] + s_2 [\phi'(z_2) - \phi'(z_3)] + \cdots \\ &+ s_{n-1} [(\phi'(z_{n-1}) - \phi'(z_n)] + s_n\phi'(z_n) , \end{aligned} \tag{4}$$

where $s_k = x_1 + x_2 + \cdots + x_k$. Since $\phi''(z) > 0$, we see that $\phi'(z_k) < \phi'(z_{k-1})$. Since $s_k(x) \geqq s_k(y)$, $k = 1, 2, \ldots, n-1$, and $s_n(x) = s_n(y)$, we see that the inequality is valid.

§ 31. An Inequality of Ostrowski

To establish the corresponding result for more general functions, let us consider a function $F(x_1, x_2, \ldots, x_n)$ that is strictly convex, so that we may write

$$F(x_1, x_2, \ldots, x_n) = \max_z \left[F(z_1, z_2, \ldots, z_n) + \sum_{i=1}^{n} (x_i - z_i) \frac{\partial F}{\partial z_i} \right]. \tag{1}$$

Once again, the maximum is taken over the region

$$z_1 \leqq z_2 \leqq \cdots \leqq z_n , \tag{2}$$

since we are interested only is values of the x_i of the same monotone nature. In order to carry through the foregoing argument, we must have certain inequalities connecting $\partial F/\partial z_i$ and $\partial F/\partial z_j$.

In particular, we would like to have

$$(z_i - z_j) \left(\frac{\partial F}{\partial z_i} - \frac{\partial F}{\partial z_j} \right) \geqq 0 \tag{3}$$

whenever $z_i \geqq z_j$. This is the natural extension of the condition that $\phi'(z_i) > \phi'(z_j)$ for $z_i > z_j$. If F satisfies this condition, we say that it satisfies a Schur condition; see Ostrowski [28.1] and Schur [29.1].

Having imposed this condition, we easily see that the foregoing proof extends without difficulty to yield the following result.

Theorem 15. *If F satisfies the Schur condition, and the $\{x_i, y_i\}$ satisfy the foregoing conditions* (28.1), *then*

$$F(x_1, x_2, \ldots, x_n) \geqq F(y_1, y_2, \ldots, y_n) . \tag{4}$$

For an interesting proof of Theorem 15 by different means, see L. Fuchs [1]. Some closely related results are due to Tatarkiewicz and Beesack; see Beesack [4]. See also Hardy, Littlewood, and Pólya [2], and Ruderman [3].

§ 32. Continuous Versions

Continuous versions of the results of §§ 28—31 have been obtained by Fan and Lorentz [1]. Write

$$f \prec g \tag{1}$$

if

$$\int_0^x f(t)\,dt \leq \int_0^x g(t)\,dt\,, 0 \leq x \leq 1\,, \tag{2}$$

and

$$\int_0^1 f(t)\,dt = \int_0^1 g(t)\,dt\,. \tag{3}$$

Then the paper [1] cited above gives necessary and sufficient conditions on the function $\phi\,(t, u_1, u_2, \ldots, u_n)$ in order that

$$\int_0^1 \phi\,(t, f_1, f_2, \ldots, f_n)\,dt \leq \int_0^1 \phi\,(t, g_1, g_2, \ldots, g_n)\,dt \tag{4}$$

for every set of decreasing bounded functions f_i and g_i with $f_i \prec g_i$, $i = 1, 2, \ldots, n$.

§ 33. Symmetric Functions

Let us now turn to some interesting inequalities of Marcus and Lopes [1]. Let $x_1, x_2, \ldots, x_n$ be a set of nonnegative quantities and denote by $E_r(x)$ the r-th elementary symmetric function of these quantities, for $r = 1, 2, \ldots, n$, i. e.,

$$\begin{aligned} E_1(x) &= x_1 + x_2 + \cdots + x_n, \\ E_2(x) &= \sum_{i \neq j} x_i x_j, \\ &\vdots \\ E_n(x) &= x_1\, x_2 \ldots x_n\,. \end{aligned} \tag{1}$$

Let $E_0(x) = 1$. Finally, write $(x) \sim (y)$ if there exists a quantity λ such that

$$x_i = \lambda y_i\,, \quad i = 1, 2, \ldots, n\,. \tag{2}$$

The first result we wish to demonstrate is the following:

Theorem 16. *For $r = 1, 2, \ldots, n$, $x_i, y_i \geq 0$, not all x_i or $y_i = 0$,*

$$\frac{E_r(x+y)}{E_{r-1}(x+y)} \geq \frac{E_r(x)}{E_{r-1}(x)} + \frac{E_r(y)}{E_{r-1}(y)}\,. \tag{3}$$

There is strict inequality unless $r = 1$, or $(x) \sim (y)$, provided that at least r of the x_i and y_i are positive.

Proof. For $r = 2$, the result follows from the identity

$$\frac{E_2(x+y)}{E_1(x+y)} - \frac{E_2(x)}{E_1(x)} - \frac{E_2(y)}{E_1(y)} = \frac{\sum\limits_{i=1}^{n}\left(x_i \sum\limits_{j=1}^{n} y_j - y_i \sum\limits_{j=1}^{n} x_j\right)^2}{2E_1(x+y)\,E_1(x)\,E_1(y)}\,. \tag{4}$$

Let us then assume that $r > 2$ and that $(x) \nsim (y)$. Write (x_j') to denote the set of $(n-1)$ quantities $x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n$. We have the two results

$$\begin{gathered}\sum_{i=1}^{n} x_i E_{r-1}(x_i') = r E_r(x)\,,\\ x_i E_{r-1}(x_i') + E_r(x_i') = E_r(x)\,.\end{gathered} \tag{5}$$

We sum the second result in (5) on i to get

$$n E_r(x) = \sum_{i=1}^{n} x_i E_{r-1}(x_i') + \sum_{i=1}^{n} E_r(x_i')\,, \tag{6}$$

and thus, using the first result in (5), we obtain

$$\sum_{i=1}^{n} E_r(x_i') = (n-r)\, E_r(x)\,. \tag{7}$$

Since

$$\begin{aligned} E_r(x) - E_r(x_i') &= x_i E_{r-1}(x_i')\\ &= x_i E_{r-1}(x) - x_i^2 E_{r-2}(x_i')\,,\end{aligned} \tag{8}$$

we obtain

$$r E_r(x) = \sum_{i=1}^{n} x_i E_{r-1}(x) - \sum_{i=1}^{n} x_i^2 E_{r-2}(x_i')\,, \tag{9}$$

and therefore

$$\begin{aligned}\frac{E_r(x)}{E_{r-1}(x)} &= \frac{1}{r}\left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \frac{x_i^2 E_{r-2}(x_i')}{E_{r-1}(x)}\right]\\ &= \frac{1}{r}\left[\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \frac{x_i^2}{x_i + E_{r-1}(x_i')/E_{r-2}(x_i')}\right].\end{aligned} \tag{10}$$

From this it follows that for $r = 1, 2, \ldots, n$, and for all x, we have

$$\begin{aligned} d(x,y) &= \frac{E_r(x+y)}{E_{r-1}(x+y)} - \frac{E_r(x)}{E_{r-1}(x)} - \frac{E_r(y)}{E_{r-1}(y)}\\ &= \frac{1}{r}\sum_{i=1}^{n}\left[\frac{x_i^2}{x_i + F_{r-1}(x_i')} + \frac{y_i^2}{y_i + F_{r-1}(y_i')}\right.\\ &\qquad \left. - \frac{(x_i+y_i)^2}{x_i + y_i + F_{r-1}(x_i'+y_i')}\right],\end{aligned} \tag{11}$$

where $F_r(x) = E_r(x)/E_{r-1}(x)$.

Assume now that the result of the theorem is valid for $r-1$; that is, assume that

$$F_{r-1}(x_i' + y_i') > F_{r-1}(x_i') + F_{r-1}(y_i')\,, \tag{12}$$

unless $(x_i') \sim (y_i')$ in which case the sign of equality holds.

Then, provided that $(x_i') \nsim (y_i')$ for some i, we have

$$d(x,y) > \frac{1}{r}\sum_{i=1}^{n}\left[\frac{x_i^2}{x_i + F_{r-1}(x_i')} + \frac{y_i^2}{y_i + F_{r-1}(y_i')} - \frac{(x_i+y_i)^2}{x_i+y_i+F_{r-1}(x_i')+F_{r-1}(y_i')}\right] \tag{13}$$

$$> \frac{1}{r}\sum_{i=1}^{n}\frac{[x_i F_{r-1}(y_i') - y_i F_{r-1}(x_i')]^2}{[x_i+F_{r-1}(x_i')]\,[y_i+F_{r-1}(y_i')]\,[x_i+y_i+F_{r-1}(x_i')+F_{r-1}(y_i')]}\,.$$

Thus the inequality holds for r provided that $(x_i') \nsim (y_i')$ for some i. If $(x_i') \sim (y_i')$ for each i, then $(x_i') = \lambda_i(y_i')$ for each i. Hence

$$[x_i F_{r-1}(y_i') - y_i F_{r-1}(x_i')]^2 = (x_i - \lambda_i y_i)^2 F_{r-1}^2(y_i')\,. \tag{14}$$

But $F_{r-1}(y_i') \neq 0$, since, by hypothesis, at least r of the y_i are positive. Hence, although $(x_i') = \lambda_i(y_i')$ implies that the first inequality in (13) is actually an equality, (14) implies that the right-hand side of (13) is positive unless $x_i = \lambda_i y_i$, in which case $(x_i) \sim (\lambda_i y_i)$, a contradiction.

§ 34. A Further Inequality

A further result concerning symmetric functions is the following:

Theorem 17. *If $\{x\}$ and $\{y\}$ are sets of n positive variables, we have*

$$[E_r(x+y)]^{1/r} \geqq [E_r(x)]^{1/r} + [E_r(y)]^{1/r}. \tag{1}$$

A stronger result, containing a discussion of equality, is contained in the paper of Lopes and Marcus [33.1] refered to above. We shall content ourselves here with the stated result.

The proof rests on a combination of Theorems **7** and **16**. Namely, by these two theorems we have

$$[E_r(x+y)]^{1/r} = \left[\frac{E_r(x+y)}{E_{r-1}(x+y)}\,\frac{E_{r-1}(x+y)}{E_{r-2}(x+y)}\cdots\frac{E_1(x+y)}{1}\right]^{1/r}$$

$$\geqq \left\{\left[\frac{E_r(x)}{E_{r-1}(x)} + \frac{E_r(y)}{E_{r-1}(y)}\right]\left[\frac{E_{r-1}(x)}{E_{r-2}(x)} + \frac{E_{r-1}(y)}{E_{r-2}(y)}\right]\cdots[E_1(x)+E_1(y)]\right\}^{1/r} \tag{2}$$

$$\geqq \left[\prod_{i=1}^{r}\frac{E_i(x)}{E_{i-1}(x)}\right]^{1/r} + \left[\prod_{i=1}^{r}\frac{E_i(y)}{E_{i-1}(y)}\right]^{1/r} = [E_r(x)]^{1/r} + [E_r(y)]^{1/r}.$$

§ 35. Some Results of Whiteley

Similar and further results are obtained by Whiteley [1,2]. Write

$$\sum_{n=0}^{\infty} T^{(n)}(a)\,t^n = \prod_{j=1}^{m}(1+a_j t)^k,\ k>0\,,$$
$$= \sum_{j=1}^{m}(1-a_j t)^k,\ k<0\,, \tag{1}$$

where the a_j are positive. Then for $k < 0$ we have

$$[T^{(n)}(a+b)]^{1/n} \leqq [T^{(n)}(a)]^{1/n} + [T^{(n)}(b)]^{1/n}, \tag{2}$$

and for $k > 0$ we have the inequality of Theorem 17.

The new result is readily obtained by means of a type of integral-representation technique we shall exploit in the next chapter. For $|t|$ small and $k > 0$, write

$$\frac{1}{(1-a_j t)^k} = \frac{1}{\Gamma(k)} \int_0^\infty e^{-s(1-a_j t)} s^{k-1} ds, \tag{3}$$

whence

$$\prod_{j=1}^{m} (1-a_j t)^k = \frac{1}{[\Gamma(k)]^m} \int_0^\infty \cdots \int_0^\infty e^{-\sum_{j=1}^m s_j} e^{\left(\sum_{j=1}^m s_j a_j\right)t} \prod_{j=1}^m s_j^{k-1} ds_1\, ds_2 \ldots ds_m. \tag{4}$$

It follows that

$$T^{(n)}(a) = \frac{1}{[\Gamma(k)]^m} \frac{1}{n!} \int_0^\infty \cdots \int_0^\infty \left(\sum_{j=1}^m s_j a_j\right)^n \phi(s)\, ds, \tag{5}$$

where

$$\phi(s)\, ds = e^{-\sum_{j=1}^m s_j} \prod_{j=1}^m s_j^{k-1}\, ds_1\, ds_2 \ldots ds_m. \tag{6}$$

The inequality of (2) is now a consequence of the Minkowski inequality.

§ 36. Hyperbolic Polynomials

Let us now discuss an important concept introduced by GÅRDING [1], that of a *hyperbolic polynomial.* Let $P(x_1, x_2, \ldots, x_n) = P(x)$ be a homogeneous polynomial of degree m in the x_i and let $a = (a_1, a_2, \ldots, a_n)$ be a set of real quantities. If the equation in s,

$$P(sa + x) = P(sa_1 + x_1, sa_2 + x_2, \ldots, sa_n + x_n) = 0,$$

has m real zeros for *all* real x_i, we say that $P(x)$ is hyperbolic with respect to a. An equivalent definition is that

$$P(sa+x) = P(a) \prod_{k=1}^{m} [s + \lambda_k(a, x)], \tag{1}$$

where $P(a) \neq 0$ and the $\lambda_k(a, x)$ are real whenever x is real.

As we shall see, this concept covers a number of results of this chapter and is also significant in connection with results we shall discuss in the chapter on matrix theory; see § 2.45.

Starting with one hyperbolic polynomial, we can form new hyperbolic polynomials by means of the following result:

Lemma 2. *If $P(x)$ is hyperbolic with respect to a, and $m > 1$, then*

$$Q(x) = \sum_{k=1}^{m} a_k \frac{\partial}{\partial x_k} P(x) \tag{2}$$

is also hyperbolic with respect to a.

The proof follows from ROLLE's theorem as in § 12. Repeated application of this lemma shows that the polynomials $\{P_k\}$ defined by

$$P(sa + x) = \sum_{k=1}^{m} s^k P_k(x) \tag{3}$$

are hyperbolic.

§ 37. Gårding's Inequality

Let $M(x^1, x^2, \ldots, x^m)$, where each of the x^i is an m-dimensional vector $x^i = (x_1^i, x_2^i, \ldots, x_m^i)$, be the completely polarized form of the polynomial $P(x)$,

$$M(x^1, x^2, \ldots, x^m) = \frac{1}{m!} \prod_{k=1}^{m} \left(\sum_{j=1}^{m} x_j^k \frac{\partial}{\partial x_j} \right) P(x) . \tag{1}$$

Theorem 18. *Let $P(x)$ be hyperbolic with respect to a, with $P(a) > 0$, $m > 1$, and let M be the completely polarized form of P. Let $x = (x_1, x_2, \ldots, x_n)$ be a set of x_i such that $P(ta + x) \neq 0$ when $t \geqq 0$. Then*

$$M(x^1, x^2, \ldots, x^m) \geqq [P(x^1)]^{1/m} \ldots [P(x^m)]^{1/m}. \tag{2}$$

Since the proof and discussion of the case of equality require a detailed analysis, we refer the reader to GÅRDING's paper [36.1].

§ 38. Examples

Two interesting examples of hyperbolic polynomials are

$$P(x) = x_1^2 - x_2^2 - \cdots - x_n^2 , \tag{1}$$

hyperbolic with respect to a whenever $P(a) > 0$, a quadratic form we shall treat in the following § 39, and

$$P(x) = x_1 x_2 \ldots x_n , \tag{2}$$

hyperbolic with respect to a for $P(a) \neq 0$. Since for the function (2) we have

$$P(sa + x) = \sum_{j=0}^{n} s^{n-j} E_j(x) , \tag{3}$$

when $a = (1, 1, \ldots, 1)$, if follows that the $E_j(x)$, the elementary symmetric functions, are hyperbolic with respect to $(1, 1, \ldots, 1)$.

As Gårding pointed out, if $P(x)$ is hyperbolic with respect to a, then the hypersurfaces $P(x) = c$ are convex for x in the set of x_i such that $P(ta + x) \neq 0$ when $t \geqq 0$. From this, some of the results of §§ 33, 34 can be derived.

Further examples will be given in Chapter 2 in connection with positive definite matrices, and in Chapter 3 in connection with the matrices introduced by LAX. Some applications of GÅRDING's inequality to the field of differential geometry are given in CHERN [2] and in GÅRDING [1].

§ 39. Lorentz Spaces

In Chapter 2, we shall restrict our attention to positive definite forms. Let us now show how certain results can be obtained for indefinite forms. We wish to demonstrate the following result:

Theorem 19. *Let*

$$\phi(x) = (x_1^p - x_2^p - \cdots - x_n^p)^{1/p}, \quad p > 1\,, \tag{1}$$

for x_i in the region R defined by

$$\begin{aligned} &\text{(a)} \qquad x_i \geqq 0\,, \\ &\text{(b)} \qquad x_1 > (x_2^p + x_3^p + \cdots + x_n^p)^{1/p}. \end{aligned} \tag{2}$$

Then for $x, y \in R$, we have

$$\phi\,(x+y) \geqq \phi(x) + \phi(y)\,. \tag{3}$$

Proof. We shall employ the quasi-linearization technique used above. Let us demonstrate that

$$\phi(x) = \min_{S(z)} \sum_{i=1}^{n} x_i z_i\,, \tag{4}$$

where $S(z)$ is the region defined by

$$\begin{aligned} &\text{(a)} \qquad z_1 \geqq 1,\ z_i \geqq 0\,, \\ &\text{(b)} \qquad (z_2^q + \cdots + z_n^q) \leqq z_1^q - 1\,, \quad q = \frac{p}{p-1}\,. \end{aligned} \tag{5}$$

Using HÖLDER's inequality, we have

$$\begin{aligned} \phi(x) &\geqq x_1 z_1 - \left(\sum_{k=2}^{n} x_k^p\right)^{1/p} \left(\sum_{k=2}^{n} z_k^q\right)^{1/q} \\ &\geqq x_1 z_1 - \left(\sum_{k=2}^{n} x_k^p\right)^{1/p} (z_1^q - 1)^{1/q}. \end{aligned} \tag{6}$$

Minimizing the right-hand side over z_1, we obtain (4). From this, the inequality (3) follows immediately. The proof given follows BELLMAN [1].

Further remarks concerning inequalities of this type for the case $p = 2$ will be found in MURNAGHAN [2].

These results can be obtained in another fashion using a representation due to BOCHNER [3]. They will be discussed again in the following Chapter 2; see §§ 2.15 and 2.16.

The quadratic form $x_1^2 + x_2^2 + x_3^2 - x_4^2$ plays a vital role in relativity theory (see SYNGE [4]), and, interestingly enough, the inequality (3) enters into the famous "twin paradox."

For a treatment of these inequalities from the standpoint of non-euclidean geometry, see ACZEL and VARGA [5].

§ 40. Converses of Inequalities

In the foregoing §§ 17, 18, we discussed the Buniakowsky-Schwarz inequality and its extension, the Hölder inequality. It is clear that no inequality of the form

$$\left(\int_0^1 uv\,dx\right)^2 \geqq k\left(\int_0^1 u^2\,dx\right)\left(\int_0^1 v^2\,dx\right) \tag{1}$$

can hold for *all* u and v in $L^2(0, 1)$ with a positive absolute constant k. It is true, however, that inequalities of this type can hold if u and v belong to certain subspaces of $L^2(0, 1)$.

Problems of this type have been discussed by FRANK and PICK [1], BLASCHKE and PICK [2], BÜCKNER [3] (cf. HARDY, LITTLEWOOD, PÓLYA [1.1]), FAVARD [4], BERWALD [5], KNESER [6], and BELLMAN [7]. Although these problems may be treated in a systematic fashion by means of the general theory of moment spaces and convex sets (cf. the treatment in DRESHER [24.1]), and greatly extended using the methods of FAVARD and BERWALD, we shall follow the treatment in BELLMAN [7], which permits a different type of generalization at the expense of less precise results.

Our first result is the following:

Theorem 20. *Let $u(x)$ and $v(x)$ be concave functions of x for $0 \leqq x \leqq 1$, normalized by the conditions*

$$\begin{aligned} &\text{(a)} \qquad \int_0^1 u^2\,dx = 1\,, \quad \int_0^1 v^2\,dx = 1\,, \\ &\text{(b)} \qquad u(0) = u(1) = 0\,, \\ &\text{(c)} \qquad v(0) = v(1) = 0\,. \end{aligned} \tag{2}$$

Then

$$\int_0^1 uv\,dx \geqq \frac{1}{2}\,. \tag{3}$$

The minimum value $\frac{1}{2}$ is attained for

$$\begin{aligned} u(x) &= x\sqrt{3},\ 0 \leqq x < 1,\ u(1) = 0\,, \\ v(x) &= (1-x)\sqrt{3},\ 0 < x \leqq 1,\ v(0) = 0\,. \end{aligned} \tag{4}$$

Proof. Let $u(x)$ belong to the class of functions over $[0,1]$ that are concave and zero at the endpoints. Consider the subclass of functions that possess nonpositive second derivatives:

$$\begin{aligned} u''(x) &= -f(x), f(x) \geqq 0, \\ u(0) &= u(1) = 0. \end{aligned} \tag{5}$$

Minimization over all functions in the larger class is equivalent to the determination of the infinum over all concave functions satisfying (5).

Let $K(x, y)$ be the GREEN's function for the operator u'' with the above boundary conditions, namely

$$\begin{aligned} K(x, y) &= x(1-y), \ 0 \leqq x \leqq y \leqq 1, \\ &= (1-x)\, y, \ 1 \geqq x \geqq y \geqq 0. \end{aligned} \tag{6}$$

Then $u(x)$ may be written

$$u(x) = \int_0^1 K(x, y) f(y)\, dy. \tag{7}$$

Using this representation, let us determine the minimum of the linear functional

$$L(u) = \int_0^1 u(x)\, h(x)\, dx \tag{8}$$

over all concave functions $u(x)$, normalized as above, where $h(x)$ is a given nonnegative function. We have

$$\begin{aligned} L(u) &= \int_0^1 h(x) \left[\int_0^1 K(x, y) f(y)\, dy\right] dx \\ &= \int_0^1 f(y) \left[\int_0^1 h(x)\, K(x, y)\, dx\right] dy, \end{aligned} \tag{9}$$

and we wish to determine the infinum over all $f \geqq 0$ for which

$$\int_0^1 u^2\, dx = \int_0^1 \int_0^1 K_2(y, z) f(y) f(z)\, dy\, dz = 1, \tag{10}$$

where

$$K_2(y, z) = \int_0^1 K(x, y)\, K(x, z)\, dx. \tag{11}$$

Reasoning by analogy with the finite-dimensional analogue, we easily see that the infinum over $f \geqq 0$, the minimum over u, is given by

$$\min_u L(u) = \min_y \left[\frac{\int_0^1 K(x, y)\, h(x)\, dx}{\sqrt{K_2(y, y)}}\right]. \tag{12}$$

We shall not give the proof here, since we shall give a proof valid for the L^p case in the following § 41.

Once we have obtained this result, Theorem 20 follows readily. We have

$$\int_0^1 u(x)\, v(x)\, dx \geqq \min_y \left[\frac{\int_0^1 K(x, y)\, v(x)\, dx}{\sqrt{K_2(y, y)}} \right]. \tag{13}$$

Applying the same inequality, this time for the nonnegative function $K(x, y)$ and the concave function $v(x)$, we obtain the end result that

$$\int_0^1 u(x)\, v(x)\, dx \geqq \min_y \left\{ \min_z \left[\frac{\int_0^1 K(x, y)\, K(x, z)\, dx}{\sqrt{K_2(y, y)}\ \sqrt{K_2(z, z)}} \right] \right\}. \tag{14}$$

It remains to show that this inequality is nontrival. A direct calculation, carried out by O. Gross, shows that the minimum is attained at $y = 1$, $z = 0$ and at the symmetric point $y = 0$, $z = 1$, yielding the value $1/2$. The minimal functions are as stated.

§ 41. L^p Case

Let us now prove the L^p generalization of Theorem 20; see [40.7]. The proof we give is due to Weinberger.

Theorem 21. *Let $u(x)$ and $v(x)$ be concave functions, normalized by the conditions*

$$\begin{aligned} &\text{(a)} \qquad \int_0^1 [u(x)]^p\, dx = 1, \int_0^1 [v(x)]^q\, dx = 1, \infty > p > 1\,, \\ &\text{(b)} \qquad u(0) = v(0) = 0, \quad u(1) = v(1) = 0\,. \end{aligned} \tag{1}$$

Then

$$\int_0^1 u(x)\, v(x)\, dx \geqq \frac{(p+1)^{1/p}\, (q+1)^{1/q}}{6}. \tag{2}$$

Note that the maximum of the right-hand side is attained at $p = q = 2$.

Proof. As above, let

$$\begin{aligned} u(x) &= \int_0^1 K(x, y)\, u_1(y)\, dy\,, \\ v(x) &= \int_0^1 K(x, y)\, v_1(y)\, dy; \end{aligned} \tag{3}$$

then

$$\begin{aligned} &\int_0^1 u(x)\, v(x)\, dx = \int\int\int K(x, y)\, K(x, z)\, u_1(y)\, v_1(z)\, dx\, dy\, dz \\ &\geqq \left[\min_{y,z} k(y, z)\right] \int\int \{\int [K(x,y)]^p\, dx\}^{1/p}\, \{\int [K(x, z)]^q dx\}\, u_1(y)\, v_1(z)\, dy\, dz, \end{aligned} \tag{4}$$

where

$$k(y, z) = \frac{\int K(x, y)\, K(x, z)\, dx}{\{\int [K(x, y)]^p dx\}^{1/p} \{\int [K(x, z)]^q dx\}^{1/q}} . \tag{5}$$

All integrals are over the interval [0,1]. Applying HÖLDER's inequality, we get

$$\begin{aligned} 1 &= \int_0^1 [u(x)]^p\, dx = \int [\int K(x, y)\, u_1(y)\, dy]\, u(x)^{p-1}\, dx \\ &= \int u_1(y) \{\int K(x, y)\, [u(x)]^{p-1}\, dx\}\, dy \\ &\leqq \int u_1(y) \{\int [K(x, y)]^p\, dx\}^{1/p} \{\int [u(x)]^p\, dx\}^{1/q}\, dy \\ &\leqq \int u_1(y) \{\int [K(x, y)]^p\, dx\}^{1/p}\, dy . \end{aligned} \tag{6}$$

Similarly we have

$$1 \leqq \int v_1(z) \{\int [K(x, z)]^q\, dx\}^{1/q}\, dz . \tag{7}$$

Combining these inequalities with (4), we see that

$$\int_0^1 u(x)\, v(x)\, dx \geqq \min_{y, z} k(y, z) . \tag{8}$$

A direct calculation, again carried out by O. GROSS, shows that the minimum has the value shown in the right-hand member of (2).

§ 42. Multidimensional Case

The same argument of §§ 40, 41 yields the following general result:

Theorem 22. *Let $u(P)$ and $v(P)$ be defined for points P belonging to a region R and satisfy the following conditions there:*

$$\begin{aligned} &\text{(a)} \qquad \Delta u, \Delta v \leqq 0, \quad P \in R, \\ &\text{(b)} \qquad u, v = 0 \ \textit{on}\ B, \textit{ the boundary of } R, \\ &\text{(c)} \qquad \int_R u^2\, dV = \int_R v^2\, dV = 1 . \end{aligned} \tag{1}$$

Let $K(P, Q)$ be the GREEN's *function for the region R, and*

$$K_2(P, Q) = \int_R K(P, P')\, K(P', Q)\, dV'. \tag{2}$$

Then

$$\int_B uv\, dV \geqq \min_{P, Q \in R} \frac{K_2(P, Q)}{\sqrt{K_2(P, P)}\, \sqrt{K_2(Q, Q)}} . \tag{3}$$

It does not seem easy to determine whether or not this is a nontrivial result, and to determine the value of the righthand side for various types of regions.

§ 43. Generalizations of Favard-Berwald

Extensive generalizations of the Frank-Pick result of § 40 can be obtained as special cases of inequalities due to FAVARD [40.4] and BERWALD [40.5].

The result of FAVARD is the following:

Theorem 23. *Let $f(x)$ be a nonnegative, continuous, concave function in $[a, b]$, not identically zero, and set*

$$\bar{f} = \frac{1}{b-a}\int_a^b f(x)\,dx. \tag{1}$$

Let $m(y)$ be a bounded and nondecreasing function in $0 \leqq y \leqq 2\bar{f}$, and set

$$\psi(y) = \int_0^y m(t)\,dt,$$

for $0 \leqq y \leqq 2\bar{f}$. Then

$$\frac{1}{2\bar{f}}\int_0^{2\bar{f}} \psi(y)\,dy \geqq \frac{1}{b-a}\int_a^b \psi(f(x))\,dx. \tag{2}$$

Favard gave corresponding results for functions of several variables, and also gave the conditions for equality in (2).

The generalization of this result due to BERWALD [40.5] follows:

Theorem 24. *Let $f(x)$ be a nonnegative, continuous, concave function that is not identically zero, for $a \leqq x \leqq b$. On $0 \leqq y \leqq y_0$, where y_0 is sufficiently large, let $\phi(y)$ be strictly monotone and continuous. Then the equation*

$$\frac{1}{z}\int_0^z \phi(y)\,dy = \frac{1}{b-a}\int_b^b \phi(f(x))\,dx \tag{3}$$

has exactly one positive root, $z = \bar{z}$.

Let $m(y)$ be bounded and monotone for $0 \leqq y \leqq \bar{z}$, and set

$$\psi(y) = \int_0^y m(t)\,d\phi(t), \tag{4}$$

a Stieltjes integral. Then

$$\frac{1}{\bar{z}}\int_0^{\bar{z}} \psi(y)\,dy \geqq \frac{1}{b-a}\int_a^b \psi(f(x))\,dx \tag{5}$$

if $\phi(y)$ and $m(y)$ are monotone in the same sense, and

$$\frac{1}{\bar{z}}\int_0^{\bar{z}} \psi(y)\,dy \leqq \frac{1}{b-a}\int_a^b \psi(f(x))\,dx \tag{6}$$

if they are monotone in opposite senses.

BERWALD also determined the cases of possible equality.

As a first application of Theorem 24, take

$$\phi(y) = y^\alpha,\ m(y) = \beta y^{\beta-\alpha}/\alpha,\ \psi(y) = y^\beta,\ 0 < \alpha < \beta\,. \tag{7}$$

Then

$$\frac{1}{z}\int_0^z \phi(y)\,dy = \frac{z^\alpha}{\alpha+1},\quad \frac{1}{\bar z}\int_0^{\bar z} \psi(y)\,dy = \frac{\bar z^\beta}{\beta+1}, \tag{8}$$

where

$$\bar z = \left[\frac{(\alpha+1)}{(b-a)}\int_a^b f(x)^\alpha\,dx\right]^{1/\alpha}. \tag{9}$$

Hence (5) yields

$$\left\{\frac{(\beta+1)}{(b-a)}\int_a^b [f(x)]^\beta\,dx\right\}^{1/\beta} \leqq \left\{\frac{(\alpha+1)}{(b-a)}\int_a^b [f(x)]^\alpha\,dx\right\}^{1/\alpha}. \tag{10}$$

For $\alpha = 1, \beta = 2$, we have the Frank-Pick inequality, while $\alpha = 1$, $\beta = p > 1$ yields the inequality of FAVARD:

$$\frac{1}{(b-a)}\int_a^b [f(x)]^p\,dx \leqq \frac{2^p}{(p+1)}\left[\frac{1}{(b-a)}\int_a^b f(x)\,dx\right]^p. \tag{11}$$

Further interesting results are

$$\frac{1}{(b-a)}\int_a^b f(x)\,dx \geqq \frac{1}{2}\max_{a\leqq x\leqq b} f(x)\,, \tag{12}$$

due to FAVARD, and

$$e^{\frac{1}{(a-b)}\int_a^b \log f(x)\,dx} \geqq \frac{2}{e}\,\frac{1}{(b-a)}\int_a^b f(x)\,dx\,, \tag{13}$$

due to BERWALD, both under the foregoing assumptions concerning $f(x)$.

Many further results, including multidimensional versions, are contained in the papers by FAVARD and BERWALD cited above.

§ 44. Other Converses of the Cauchy Theorem

In the preceding §§ 40—43, we considered one type of converse of CAUCHY's inequality. Let us now state another, due to P. SCHWEITZER and PÓLYA-SZEGÖ [1], where a proof may be found.

Theorem 25. *If* $0 < m_1 \leqq x_i \leqq M_1$, $0 < m_2 \leqq y_i \leqq M_2$, *then*

$$1 \leqq \frac{\left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n y_i^2\right)}{\left(\sum_{i=1}^n x_i y_i\right)^2} \leqq \left[\frac{\left(\frac{M_1 M_2}{m_1 m_2}\right)^{1/2} + \left(\frac{m_1 m_2}{M_1 M_2}\right)^{1/2}}{2}\right]^2. \tag{1}$$

The analogous continuous version also holds.

As in shown by DRESHER [24.1], the classical inequalities of CAUCHY and HÖLDER, and improvements, can be obtained as consequences of the theory of moment spaces, along the lines we shall discuss in Chapter 3. The preceding result is of the same nature and can be derived by these methods. Extensions to the multidimensional case have been given by MADANSKY [2].

An extension of this result is contained in a paper by G. S. WATSON [3], with a proof by J. W. S. CASSELS; see also W. GREUB and W. RHEINBOLDT [4].

Theorem 26. *Let* $x_i, y_i > 0$, $w_i \geqq 0$, $i = 1, 2, \ldots, n$, *where not all* $w_i = 0$. *Then*

$$1 \leqq \frac{\left(\sum_{i=1}^{n} x_i^2 w_i\right)\left(\sum_{i=1}^{n} y_i^2 w_i\right)}{\left(\sum_{i=1}^{n} x_i y_i w_i\right)^2} \leqq \max_{i,j} \frac{(x_i y_j + x_j y_i)^2}{4 x_i x_j y_i y_j}. \tag{2}$$

A further result in this direction is due to KY FAN and J. TODD [5]:

Theorem 27. *Let* a_i, b_i, $i = 1, 2, \ldots, n$, *be real numbers with each* $a_j b_i - a_i b_j \neq 0$ $(i \neq j)$. *Then*

$$\frac{\binom{n}{2}^2 \Sigma a_i^2}{\Sigma a_i^2 \, \Sigma b_i^2 - (\Sigma a_i b_i)^2} \leqq \sum_{i=1}^{n} \left(\sum_{j \neq i} \frac{a_j}{a_j b_i - a_i b_j}\right)^2.$$

This is an extension of an earlier result of J. B. CHASSAN [6].

§ 45. Refinements of the Cauchy-Buniakowsky-Schwarz Inequalities

Having established the nonnegativity of the functional

$$I(u, v) = \left(\int_0^1 u^2\, dt\right)\left(\int_0^1 v^2\, dt\right) - \left(\int_0^1 uv\, dt\right)^2, \tag{1}$$

we naturally are interested in obtaining a more precise lower bound than zero. We can do this whenever the functions or functionals under consideration are quadratic in the following fashion.

Reverting to inner products, consider the function

$$J(u, v) = (u, u)(v, v) - (u, v)^2, \tag{2}$$

assumed nonnegative for all u and v.

Replace u by $rf + sg$, where r and s are scalars. Then $J(u, v) \geqq 0$ yields the inequality

$$\begin{aligned} r^2 [(f,f)(v,v) - (f,v)^2] + 2rs\, [(f,g)(v,v) - (f,v)(g,v)] \\ + s^2 [(g,g)(v,v) - (g,v)^2] \geqq 0, \end{aligned} \tag{3}$$

which, in turn, leads to

$$\begin{aligned}&[(f,f)\,(v,v)-(f,v)^2]\,[(g,g)\,(v,v)-(g,v)^2]\\ &\geqq [(f,g)\,(v,v)-(f,v)\,(g,v)]^2.\end{aligned}\tag{4}$$

This inequality asserts that a new inner product, defined for fixed v and any two elements f and g,

$$[f,g;v]=(f,g)\,(v,v)-(f,v)\,(g,v)\tag{5}$$

satisfies the same conditions as (f,g), namely

$$[f,f;v]\,[g,g;v]\geqq [f,g;v]^2.\tag{6}$$

Consequently, we can repeat the foregoing procedure, and obtain in this way a chain of inequalities, each stronger than the preceding. These inequalities are equivalent to those given in (§ 5) of Chapter 2.

§ 46. A Result of Mohr and Noll

The following result was established by MOHR and NOLL [1], a refinement of the inequality $\left(\int_a^b f\,dt\right)^2 \leqq (b-a)\int_a^b f^2\,dt$.

Theorem 28. *Let $f(t)$ possess a continuous n-th derivative in $[a,b]$. Then*

$$\begin{aligned}\left(\int_a^b f(t)\,dt\right)^2 &= (b-a)\sum_{k=0}^{n=1}\frac{(-1)^k}{k!\,(k+1)!}\\ &\quad\cdot\int_a^b [f^{(k)}(t)]^2\,[(b-t)\,(t-a)]^k\,dt+(-1)^n R_n,\end{aligned}\tag{1}$$

where

$$R_n=2\,(n!)^{-2}\iint_{a\leqq t\leqq s\leqq b} f^{(n)}(s)\,f^{(n)}(t)\,[(b-s)\,(t-s)]^n\,ds\,dt.\tag{2}$$

That R_n is nonnegative follows from the alternative representation

$$R_n=\int\cdots\int\left[\int_{x_1}^{y_1} f^{(n)}(t)\,dt\right]^2 dx_1\,dy_1\ldots dx_n\,dy_n,\tag{3}$$

where the integration is over the domain

$$a\leqq x_n\leqq\cdots\leqq x_1\leqq y_1\leqq\cdots\leqq y_n\leqq b.\tag{4}$$

More general results can be obtained by using the inverse operators associated with the linear differential equation

$$L(u)=f(t),\text{ where } L(u)=d^n u/dt^n+p_1(t)\,d^{n-1}u/dt^{n-1}+\cdots+p_n(t)\,u.$$

§ 47. Generation of New Inequalities from Old

The inequality (45.4) is a particular case of more general inequalities that can systematically be generated in the following way. Let

$$F(u,v)\geqq 0\tag{1}$$

be an inequality with the property than equality holds only for $u = v$. We now proceed to study the perturbed inequality obtained by setting

$$\begin{aligned} u &= w + rf\,, \\ v &= w + sg\,, \end{aligned} \tag{2}$$

where w, f and g are elements of the same nature as u and v, and r and s are scalar quantities. Expanding (1) about the point $r = s = 0$, we obtain a result of the form

$$\begin{aligned} A\,(f, g, w)\, r^2 + 2B\,(f, g, w)\, rs + C\,(f, g, w)\, s^2 \\ + O\,(r^3, r^2 s, rs^2, s^3) \geqq 0\,. \end{aligned} \tag{3}$$

Choosing r and s sufficiently small, we see that this implies

$$A\,(f, g, w)\, r^2 + 2B\,(f, g, w)\, rs + C\,(f, g, w)\, s^2 \geqq 0 \tag{4}$$

for all r and s, and thus that

$$B^2\,(f, g, w) \leqq A\,(f, g, w)\; C\,(f, g, w)\,. \tag{5}$$

The result (45.4) is obtained by applying the technique just sketched to the Buniakowsky-Schwarz inequality. Other interesting results are obtained from the Hölder inequality and the arithmetic-mean — geometric-mean inequality.

§ 48. Refinement of Arithmetic-mean — Geometric-mean Inequality

As we know, the inequality holds in the arithmetic-mean —geometric-mean relation unless all the x_i are equal. In connection with the theory of algebraic numbers, it is of some interest to measure the nonequality of the x_i by means of the discriminant

$$d_n = \prod_{1 \leqq i < j < n} (x_i - x_j)^2\,, \tag{1}$$

and to estimate a lower bound for the ratio

$$\frac{\left(\sum\limits_{i=1}^{n} x_i\right)^n}{x_1\, x_2\, \ldots\, x_n} \tag{2}$$

in terms of d_n. Results of this type were first given by Schur [1], and subsequent investigations were carried out by Siegel [2], Kober [3], Dinghas [4].

§ 49. Inequalities with Alternating Signs

The following result was obtained by Szegö [1].

Theorem 29. *Let* $a_1 \geqq a_2 \geqq \cdots \geqq a_{2n-1} \geqq 0$, *and let* $f(x)$ *be a convex function defined on* $[0, a_1]$. *Then*

$$\sum_{j=1}^{2n-1} (-1)^{j-1} f(a_j) \geqq f\left(\sum_{j=1}^{2n-1} (-1)^{j-1}\, a_j\right). \tag{1}$$

An analogous result was derived by BELLMAN [2] by means of a simple comparison of areas:

Theorem 30. *Let* $a_1 \geqq a_2 \geqq \cdots \geqq a_{2n-1} \geqq 0$, *and* $f(x)$ *be a convex function defined on* $[0, a_1]$. *Then, if* $f(0) \leqq 0$,

$$\sum_{j=1}^{n} (-1)^{j-1} f(a_j) \geqq f\left(\sum_{j=1}^{n} (-1)^{j-1} a_j\right). \tag{2}$$

The special case $f(x) = x^r$, $r > 1$, was established independently by WEINBERGER [3].

Using the Schur majorization of § 19, OLKIN [4] derived a more general inequality containing the two foregoing results as special cases:

Theorem 31. *Let* $1 \geqq w_1 \geqq w_2 \geqq \cdots \geqq w_n \geqq 0$, $a_1 \geqq a_2 \geqq \cdots \geqq a_n \geqq 0$, *and let* $f(x)$ *be a convex function defined on* $[0, a_1]$. *Then*

$$\left[1 - \sum_{j=1}^{n} (-1)^{j-1} w_j\right] f(0) + \sum_{j=1}^{n} (-1)^{j-1} w_j f(a_j) \geqq f\left[\sum_{j=1}^{n} (-1)^{j-1} w_j a_j\right]. \tag{3}$$

Rather than follow OLKIN's procedure, we shall establish an inequality containing the foregoing result as a special case; see §§ 50, 51.

A number of OLKIN's theorems are contained in BRUNK [6], where more general results are given. These are related to the theorems established earlier in §§ 28—32.

§ 50. Steffensen's Inequality

Let us establish the following result due to STEFFENSEN [1]. The proof follows BELLMAN [2].

Theorem 32. *Let*

$$\begin{aligned} &\text{(a)}\quad f(t) \text{ be nonnegative and monotone decreasing in } [a, b]\,, \\ &\text{(b)}\quad g(t) \text{ satisfy the constraint } 0 \leqq g(t) \leqq 1\,,\ t \text{ in } [a, b]\,. \end{aligned} \tag{1}$$

Then

$$\int_{b-c}^{b} f(t)\, dt \leqq \int_{a}^{b} f(t)\, g(t)\, dt \leqq \int_{a}^{a+c} f(t)\, dt\,, \tag{2}$$

where

$$c = \int_{a}^{b} g(t)\, dt\,. \tag{3}$$

Proof. Define the function $u(s)$ by means of the relationship

$$\int_{a}^{s} f(t)\, g(t)\, dt = \int_{a}^{u} f(t)\, dt\,. \tag{4}$$

It is easy to see, as a consequence of our assumptions concerning f and g, that $u(a) = a$, that $u(s)$ is continuous and monotone increasing as s goes from a to b, and that $u(s) \leqq s$. Upon differentiating, we have

$$f(u)\,\frac{du}{ds} = f(s)\,g(s)\,, \tag{5}$$

whence

$$\frac{du}{ds} = \frac{f(s)}{f(u)}\,g(s) \leqq g(s)\,. \tag{6}$$

Hence

$$u \leqq a + \int_a^s g(t)\,dt\,. \tag{7}$$

This yields the right-hand side of (2), and the left-hand side is derived similarly.

§ 51. Brunk-Olkin Inequality

To obtain Olkin's inequality (49.3), let $g(t)$ be the function defined as follows:

$$g(t) = \lambda_k,\ a_{k+1} \leqq t \leqq a_k,\ k = 1, 2, \ldots, n-1\,, \tag{1}$$

where

$$\lambda_1 = w_1,\ \lambda_2 = w_1 - w_2,\ \lambda_3 = w_1 - w_2 + w_3\,, \tag{2}$$

and so on, and let $h(t) = f'(t)$. This choice actually yields a slightly stronger result than that contained in Theorem 31.

§ 52. Extensions of Steffensen's Inequality

Many further results can be established in the same fashion. For example, there is the following extension of Steffensen's inequality:

Theorem 33. *Let*

(a) $f(t)$ *be nonnegative and monotone decreasing in* $[a, b]$,

(b) $f \in L^p[a, b]$, (1)

(c) $g(t) \geqq 0, \int_a^b g^q dt \leqq 1$,

where $p > 1$, $q = p/(p-1)$. *Then*

$$\left(\int_a^b fg\,dt\right)^p \leqq \int_a^{a+c} f^p dt\,, \tag{2}$$

where

$$c = a + \left(\int_a^b g\,dt\right)^p. \tag{3}$$

Bibliographical Notes

§ 1. The source book in the theory of inequalities is

1. HARDY, G. H., J. E. LITTLEWOOD and G. PÓLYA: Inequalities. London: Cambridge University Press 1951.

Historical remarks, references, and extensive generalizations will be found there.

§ 2. The generalized cosine introduced here plays an important role in mathematical statistics, where it is the correlation coefficient.

1. CAUCHY, A. L.: Cours d'analyse de l'Ecole Royale Polytechnique. I[re] partie. Analyse algebrique. Paris: Debure frères 1821. Oeuvres complètes, II[e] série, III.

§ 3. See § 6 of Chapter 2.

§ 7. For a number of other applications of this method, see

1. BELLMAN, R.: Dynamic programming. Princeton, N. J.: Princeton University Press 1957.

§ 8. This geometrical technique that we pass over so lightly can be made the basis of the theory of inequalities. We shall discuss this approach in great detail in our second volume. See HARDY-LITTLEWOOD-PÓLYA [1.1], and

1. BONNESEN, T., and W. FENCHEL: Theorie der konvexen Körper. Berlin: Julius Springer Verlag 1951.
2. LORCH, E. R.: Differentiable inequalities and the theory of convex bodies. Trans. Am. Math. Soc. **71**, 243—266 (1951).

For a general discussion of convex functions and their applications, see

3. BECKENBACH, E. F.: Convex functions. Bull. Am. Math. Soc. **54**, 439—460 (1948).
4. GREEN, J. W.: Recent applications of convex functions. Am. Math. Monthly **61**, 449—454 (1954).

§ 9.

1. BOHR, H.: The arithmetic and geometric means. J. London Math. Soc. **10**, 114 (1935).

§ 10.

1. HURWITZ, A.: Über den Vergleich des arithmetischen und des geometrischen Mittels. Math. Werke, 505—507. Basel: E. Berkhäuser 1933.
2. HURWITZ, A.: Über die Erzeugung der Invarianten durch Integration. Math. Werke, 546—564. Basel: E. Birkhäuser 1933.

§ 11. This proof was brought to our attention by KY FAN. See

1. EHLERS, G.: In: W. D. HAYES (ed): Colloquium on linear equations. Office of Naval Research Technical Report ONRL-35-54, 1954.

Another simple inductive proof is given in

2. DIANANDA, P. H.: A simple proof of the arithmetic mean-geometric mean inequality. Am. Math. Monthly **67**, 1007 (1960).

§ 12.

1. GAUSS, C. F.: Werke, **10**, 1, 172—292. Leipzig: B. G. Teubner 1917.
2. SCHAPIRA, H.: Bemerkungen zu der Grenzfunktion algebraischer Iteration. Schlömilch Z. **32**, 310—314 (1887).
3. SCHLESINGER, L.: Über GAUSS' Jugendarbeiten zum arithmetisch-geometrischen Mittel. Deutsche Math. Ver. **20**, 396—403 (1911),

where many other references will be found.

4. BELLMAN, R.: The symmetric mean. Math. Mirror, Brooklyn College **9**, 5—6 (1941).
5. MACLAURIN, C.: A second letter to Martin Folges, Esq.; concerning the roots of equations with the demonstration of other rules in algebra. Phil. Trans. **36**, 59—96 (1729).

§ 13.
1. JACOBSTHAL, E.: Über das arithmetische und geometrische Mittel. Norske Vid. Selsk. Forh. (Trondheim) **23**, 122 (1951).

§ 14. We follow HARDY-LITTLEWOOD-PÓLYA [1.1].

§ 15.
1. YOUNG, W. H.: On classes of summable functions and their Fourier series. Proc. Roy. Soc. (A) **87**, 225–229 (1912).

§ 16. The fact that $M_t(x, \alpha)$ is not necessarily a convexo-concave function was pointed out in
1. SHNIAD, H.: On the convexity of mean value functions. Bull. Am. Math. Soc. **54**, 770–776 (1948).

The inequality $S_{t_1}(x) \leqq S_{t_2}(x)$ for $0 < t_1 < t_2$ was established by JENSEN in
2. JENSEN, J. L. W. V.: Sur les fonctions convexes et les inégalités entre les valeurs moyennes. Acta Math. **30**, 175–193 (1906)

as an application of his new theory of convex functions. As Jensen stated, however, the inequality had previously been proven by PRINGSHEIM in
3. PRINGSHEIM, A.: Zur Theorie der ganzen transzendenten Funktionen (Nachträge). Münch. S.-B. **32**, 295–304 (1902).

PRINGSHEIM, in turn, attributed his proof to LÜROTH. Curiously enough, while JENSEN used convex functions in establishing the inequality, he did not show that the function $S_t(x)$ is itself convex for $t > 0$; this was not done until later, in
4. BONNESEN, T.: En bemaerkning om konvekse funktioner. Matem. Tidsskr. (B) **1928**, 18–20

for $t > 1$, and, by a different method, in
5. BECKENBACH, E. F.: An inequality of JENSEN. Am. Math. Monthly **53**, 501–505 (1946)

for $t > 0$.

§ 17.
1. HÖLDER, O.: Über einen Mittelwertsatz. Göttinger Nachr. **1889**, 38–47.
2. MINKOWSKI, H.: Geometrie der Zahlen. I. Leipzig: B. G. Teubner 1896.

§ 18.
1. BÔCHER, M.: Introduction to higher algebra. New York: Macmillan and Co. 1907.
2. OGURA, K.: Generalizations of BESSEL's and GRAM's inequalities and the elliptic space of infinitely many dimensions. Tôhoku Math. J. **18**, 1–22 (1920).

§ 19. This method has been developed independently by many different authors, with the result that it is quite difficult to assign any priority. See, however, the discussion in § 25.
1. BELLMAN, R., I. GLICKSBERG and O. GROSS: Some aspects of the mathematical theory of control processes. Santa Monica, Calif.: The RAND Corporation, Report R-313, 1958.
2. ZYGMUND, A.: On certain integrals. Trans. Am. Math. Soc. **55**, 170–204 (1944).

§ 23.
1. BECKENBACH, E. F.: A class of mean-value functions. Am. Math. Monthly **57**, 1–6 (1950).

§ 24.
1. DRESHER, M.: Moment spaces and inequalities. Duke Math. J. **20**, 261–271 (1953).
2. DANSKIN, J. M.: Dresher's inequality. Am. Math. Monthly **49**, 687–688 (1952).

§ 25.
1. MINKOWSKI, H.: See BONNESEN-FENCHEL [8.1].
2. MAHLER, K.: Ein Übertragungsprinzip für konvexe Körper. Časopis Mat. Fysik **68**, 93–102 (1939).

3. YOUNG, L. C.: On an inequality of MARCEL RIESZ. Ann. of Math. **40**, 567–574 (1939).
4. FENCHEL, W.: A remark on convex sets and polarity. Comm. Sém. Math. de Lund, Tome Supp., 1952, 82–89.
5. LORCH, E. R.: On the volume of smooth convex bodies in Hilbert space. Math. Z. **61**, 391–407 (1955).
6. — Convexity and normed spaces. Publ. Inst. Math. Acad. Sci. Serbe **4**, 109–112 (1952).
7. — Su certe extensioni del concetto di volume. Rend. Accad. Lincei (8) **16**, 25–29 (1954)

§ 26. We shall discuss this type of quasilinearization again in Chapter 4.

1. BELLMAN, R.: Functional equations in the theory of dynamic programming — V: positivity and quasi-linearity. Proc. Nat. Acad. Sci. USA **41**, 743–746 (1955).
2. — On the representation of the solution of a class of stochastic differential equations. Proc. Am. Math. Soc. **9**, 326–327 (1958).
3. KALABA, R.: On nonlinear differential equations, the maximum operation, and monotone convergence. J. Math. and Mech. **8**, 519–574 (1959).

The inequality

$$\left(\sum_{i=1}^{n} a_i x_i^2\right)\left(\sum_{i=1}^{n} \frac{1}{a_i}\right) \geqq \sum_{i=1}^{n} x_i$$

yields the quasilinearization

$$\frac{1}{\left(\sum_{i=1}^{n} 1/a_i\right)} = \min_{x_i} \sum_{i=1}^{n} a_i x_i^2,$$

where $a_i > 0$ and the minimum is taken over $x_i \geqq 0$, $\sum_{i=1}^{n} x_i = 1$.

§ 27. See

1. ZYGMUND, A.: Two notes on inequalities. J. Math. Physics **21**, 117–123 (1942).
2. MARCIENKIEWICZ, J., and A. ZYGMUND: Quelques inégalités pour les opérations linéaires. Fund Math. **32**, 115–121 (1939).

§ 28. This result plays a role in matrix theory, as will be indicated in Chapter 2. See also

1. OSTROWSKI, A.: Sur quelques applications des fonctions convexes et concaves au sens de I. SCHUR. J. Math. Pure Appl. **31**, 253–292 (1952).
2. KARAMATA, J.: Sur une inégalité relative aux fonctions convexes. Publ. Math. Univ. Belgrade **1**, 145–148 (1932).

It and the result of the next section play important parts in probability theory and mathematical statistic. See

3. BLACKWELL, D.: Comparison of experiments. Proc. Second Berkeley Symposium on Mathematical Statistics and Probability, 1950, 93–102. Berkeley, Calif.: University of California Press 1951,

where unpublished work of BLACKWELL, BOHNENBLUST, and SHERMAN is also discussed.

§ 29. See

1. SCHUR, I.: Über eine Klasse von Mittelbildungen mit Anwendungen auf die Determinantentheorie. Sitzber. Berl. Math. Ges. **22**, 9–20 (1923).
2. BIRKHOFF, G.: Tres observaciónes sobre el álgebra lineal. Rev. Univ. Nac. Tucuman (A) **5**, 147–151 (1946).
3. SCHREIBER, S.: On a result of S. SHERMAN concerning doubly stochastic matrices. Proc. Am. Math. Soc. **9**, 350–353 (1958).

4. MIRSKY, L.: Proofs of two theorems on doubly-stochastic matrices. Proc. Am. Math. Soc. **9**, 371–374 (1958).
5. RYSER, H. J.: Matrices of zeros and ones. Bull. Am. Math. Soc. **66**, 442–464 (1960).

Further comments and results may be found in HARDY-LITTLEWOOD-PÓLYA [1.1].

§ 31.
1. FUCHS, L.: A new proof of an inequality of HARDY-LITTLEWOOD-PÓLYA. Mat. Tisdkr. (B) **1947**, 53–54.
2. HARDY, G. H., J. E. LITTLEWOOD and G. PÓLYA: Some simple inequalities satisfied by convex functions. Messenger of Math. **58**, 145–152 (1929).
3. RUDERMAN, H. D.: Two new inequalities. Am. Math. Monthly **59**, 29–32 (1952).
4. BEESACK, P. R.: A note on an integral inequality. Proc. Am. Math. Soc. **8**, 875–879 (1957).

§ 32.
1. FAN, K., and G. G. LORENTZ: An integral inequality. Am. Math. Monthly **61**, 626–631 (1954).

§ 33.
1. MARCUS, M., and L. LOPES: Inequalities for symmetric functions and Hermitian matrices. Canad. J. Math. **8**, 524–531 (1956).
2. BULLEN, P., and M. MARCUS: Symmetric means and matrix inequalities. 1961, to appear.

§ 35.
1. WHITELEY, J. N.: Some inequalities concerning symmetric functions. Mathematika **5**, 49–57 (1958).

Further results, generalizing some due to NEWTON, are given in
2. WHITELEY, J. N.: A generalization of a theorem of NEWTON. Proc. Am. Math. Soc. **1961**, to appear.

§ 36.
1. GÅRDING, L.: An inequality for hyperbolic polynomials. J. Math. and Mech. **8**, 957–966 (1959).

§ 38. See an earlier paper,
1. GÅRDING, L.: Linear hyperbolic partial differential equations with constant coefficients. Acta Math. **85**, 1–62 (1951),

and
2. CHERN, S.: Integral formulas for hypersurfaces in Euclidean space and their application to uniqueness theorems. J. Math. and Mech. **8**, 947–956 (1959).

§ 39. These results have been found independently by so many authors that it is difficult to assign priority.
1. BELLMAN, R.: On an inequality concerning an indefinite form. Am. Math. Monthly **63**, 108–109 (1956).
2. MURNAGHAN, F. D.: SCHWARZ's inequality and Lorentz spaces. Proc. Nat. Acad. Sci. USA **36**, 673–676 (1950).
3. BOCHNER, S.: Group invariance of CAUCHY's formula in several variables. Ann. of Math. **45**, 686–707 (1944).
4. SYNGE, J. L.: Relativity, the special theory. New York: Interscience Publ. 1956.
5. ACZEL, J., and O. VARGA: Bemerkung zur Cayley-Kleinschen Maßbestimmung. Publ. Math., Debrecen **4**, 3–15 (1955).

§ 40.
1. FRANK, PH., and G. PICK: Distanzschätzungen im Funktionenraum. I. Math. Ann. **76**, 354–375 (1915).
2. BLASCHKE, W., and G. PICK: Distanzabschätzungen im Funktionsraum II. Math. Ann. **77**, 277–302 (1916).

3. BÜCKNER, H.: Untere Schranken für skalare Produkte von Vektoren und für analoge Integralausdrücke. Ann. Mat. Pura Appl. (4) **28**, 237–261 (1949).
4. FAVARD, J.: Sur les valeurs moyennes. Bull. Sci. Math. (2) **57**, 54–64 (1933).
5. BERWALD, L.: Verallgemeinerung eines Mittelwertsatzes von J. FAVARD, für positive konkave Funktionen. Acta Math. **79**, 17–37 (1947).
6. KNESER, H.: Bemerkung über die gemischten Inhalte in vier Dimensionen. Math. Ann. **115**, 132–135 (1937).
7. BELLMAN, R.: Converses of SCHWARZ's inequality. Duke Math. J. **23**, 429–434 (1956).

§ 44.

1. PÓLYA, G., and G. SZEGÖ: Aufgaben und Lehrsätze aus der Analysis. I, 57. Berlin: Julius Springer Verlag 1925.
2. MADANSKY, A.: Bounds on the expectation of a convex function of a multivariate random variable. Ann. Math. Stat. **30**, 743–746 (1959).
3. WATSON, G. S.: Serial correlation in regression analysis. I. Biometrika **42**, 327–341 (1955).
4. GREUB, W., and W. RHEINBOLDT: On a generalization of an inequality of L. V. KANTOROVICH. Proc. Am. Math. Soc. **10**, 407–415 (1959).
5. FAN, K., and J. TODD: A determinantal inequality. J. London Math. Soc. **30**, 58-64 (1955).
6. CHASSAN, J. B.: A statistical derivation of a pair of trigonometric inequalities. Am. Math. Monthly **62**, 353–356 (1955).

§ 46.

1. MOHR, E., and W. NOLL: Eine Bemerkung zur Schwarzschen Ungleichheit. Math. Nachr. **7**, 55–59 (1952).

§ 48.

1. SCHUR, I.: Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten. Math. Z. **1**, 377–402 (1918).
2. SIEGEL, C. L.: The trace of totally positive and real algebraic integers. Ann. of Math. **46**, 302–312 (1945).
3. KOBER, H.: On the arithmetic and geometric means and on HÖLDER's inequality. Proc. Am. Math. Soc. **9**, 452–459 (1958).
4. DINGHAS, A.: Zur Abschätzung arithmetischer Mittel reeller Zahlen durch Differenzprodukte derselben. Rend. Circ. Mat. Palermo (2) **2**, 177–202 (1953).

§ 49.

1. SZEGÖ, G.: Über eine Verallgemeinerung des Dirichletschen Integrals. Math. Z. **52**, 676–685 (1950).
2. BELLMAN, R.: On an inequality of WEINBERGER. Am. Math. Monthly **60**, 402 (1953).
3. WEINBERGER, H. F.: An inequality with alternating signs. Proc. Nat. Acad. Sci. USA **38**, 611–613 (1952).
4. OLKIN, I.: On inequalities of SZEGÖ and BELLMAN. Proc. Nat. Acad. Sci. USA **45**, 230–231 (1959).
5. WRIGHT, E. M.: An inequality for convex functions. Am. Math. Monthly **61**, 620–622 (1954).
6. BRUNK, H. D.: On an inequality for convex functions. Prcc. Am. Math. Soc. **7**, 817—824 (1956).

§ 50.

1. STEFFENSEN, J. F.: Bounds of certain trigonometric integrals. Tenth Scandinavian Math. Congress 1946, 181–186. Copenhagen: J. Gjellerups Forlag 1947.
2. BELLMAN, R.: On inequalities with alternating signs. Proc. Am. Math. Soc. **10**, 807–809 (1959).

We have omitted any discussion of statistical metric spaces, concepts introduced by WALD and MENGER, for which see

3. SCHWEIZER, B., and A. SKLAR: Statistical metric spaces. Pacific J. Math. **10**, 313–334 (1960),

and of inequalities arising in probability theory and information theory, where the entropy function $\sum_i p_i \log p_i$ plays a significant role; see

4. LEIPNIK, R.: The extended entropy uncertainty principle. Information and Control **3**, 18–25 (1960).
5. BELLMAN, R., and R. KALABA: Dynamic programming and statistical communication theory. Proc. Nat. Acad. Sci. USA **43**, 749–751 (1957).
6. MOYAL, J. E.: Quantum mechanics as a statistical theory. Proc. Cambridge Phil. Soc. **45**, 99–124 (1949. Appendix 1).

Elegant extensions of the triangle inequality in the form of statements pertaining to the positive definite character of certain quadratic forms are contained in

7. BEURLING, A.: Sur quelques formes positives avec une application à la théorie ergodique. Acta Math. **78**, 319–334 (1946).

Chapter 2

Positive Definite Matrices, Characteristic Roots, and Positive Matrices

§ 1. Introduction

As soon as we leave the well-traversed fields of real and complex numbers for the broader and relatively unexplored domains of hypercomplex numbers, we open the way for the introduction of many different types of ordering relationships. In this chapter, we shall discuss a variety of interesting inequalities centering about the theme of *matrices*. As we shall see, the basic concept of *positive number* can be extended to matrices in many different and significant ways.

The simplest and most immediate extension is the notion of *positive definite* matrix. We shall begin by focusing our attention on the characteristic roots of matrices of this type and on certain functions of the characteristic roots. The most interesting function of this sort is the product of the characteristic roots, the *determinant* of the matrix.

Of the many routes that can be followed toward our goal of obtaining properties of the determinant, we shall pursue one based on integral representations of determinants. After a few results have been derived from an elementary representation of the corresponding positive definite quadratic form as a sum of squares, some more intricate results are obtained from an integral representation of the determinant, due to INGHAM and SIEGEL, and from generalizations due to BELLMAN and OLKIN.

To get corresponding results for positive definite *hermitian* matrices, we shall derive a new identity for the determinant of a hermitian matrix.

This representation enables us to exhibit a number of results by means of the same type of argument that is valid for symmetric matrices.

Several of the results we derive are due originally to KY FAN, who obtained them by means of arguments based on variational formulas extending FISCHER's classical min-max determination of the characteristic roots. This is another powerful method that might be systematically exploited.

Turning from the study of positive definite matrices, we next consider *positive matrices*, a class of matrices introduced into analysis by PERRON and extensively cultivated by FROBENIUS. In recent years, these matrices have assumed a prominent position in the study of computational algorithms for the numerical solution of partial differential equations in the study of branching processes, in the theory of games and linear programming, and in mathematical economics.

In the case of positive matrices, we are able to obtain a variational determination of one characteristic root, the root of largest absolute value, which turns out to be real. The representation thus obtained enables us to establish an interesting result concerning the behavior of this root as a function of the matrix.

The theory of positive matrices is only the surface outcropping of two extremely rich veins, the theory of *positive operators* and the theory of *variation-diminishing transformations*. We shall have only some very brief remarks to make concerning these fields in which GANTMACHER, KREIN, SCHOENBERG, KARLIN and MCGREGOR, and others have done so much. These two theories merit treatises of their own.

In addition to the foregoing concepts of positivity, there are also the *domain of positivity* introduced by KOECHER, which is connected with the integral of INGHAM and SIEGEL, and the *positive transformations* of LOEWNER, which turn out to have great significance in the scattering theory of quantum mechanics and also in the analysis of linear electrical networks.

We shall briefly mention the recent results of Lax concerning a class of matrices of importance in the field of hyperbolic partial differential equations. The general results of GÅRDING, discussed in §§ 36—38 of Chapter 1, show the reason for the similarity that exists between the theory of the characteristic roots of these matrices and the theory of the characteristic roots of hermitian matrices.

It will be apparent from our brief sketch of some results, our hints of others, and our references to still others, that the study of the many different types of order relationships associated with finite-dimensional and infinite-dimensional operators has just begun, and that many beautiful ideas and elegant results lie ahead. In view of the great number of inequalities that exist relating to determinants and characteristic roots of matrices, our aim has been to indicate some of the many different approaches that can be used, rather than to attempt an encyclopedic account.

§ 2. Positive Definite Matrices

Let us begin by recalling some fundamental notions and exhibiting the symbolism we shall employ.

A real, square matrix, $A = (a_{ij})$, $i, j = 1, 2, \ldots, n$, is said to be *symmetric* provided $a_{ij} = a_{ji}$. A real, symmetric matrix is said to be *positive definite* provided the quadratic form

$$Q(x) = \sum_{i,j=1}^{n} a_{ij} x_i x_j \tag{1}$$

is positive for all nontrivial sets of values of the real variables x_i, i. e., for $(x_1, x_2, \ldots, x_n) \neq (0, 0, \ldots, 0)$ [1].

A complex matrix $H = (h_{ij})$, $i, j = 1, 2, \ldots, n$, is said to be *hermitian* provided the h_{ij} and h_{ji} are conjugate imaginaries, i. e., $h_{ij} = \overline{h_{ji}}$. A hermitian matrix is said to be *positive definite* if the hermitian form

$$P(z) = \sum_{i,j=1}^{n} h_{ij} z_i \overline{z_j} \tag{2}$$

is positive for all nontrivial sets of values of the complex variables z_i. Of course, the hermitian form (2) reduces to (1) when the z_i and h_{ij} are real.

If the quadratic or hermitian form is merely nonnegative, we call the corresponding matrix *positive semidefinite* or *nonnegative definite.*

Let us, as usual, write

$$(x, y) = \sum_{i=1}^{n} x_i y_i \tag{3}$$

for any two n-dimensional vectors x and y, the *inner product* of x and y. Then we may write the compact and illuminating representations

$$\begin{aligned} Q(x) &= (x, Ax)\,, \\ P(z) &= (z, H\bar{z})\,. \end{aligned} \tag{4}$$

Finally, let us write $|A|$ to denote the determinant of the matrix A. This convention will be followed throughout Chapter 2; when vertical bars denote absolute value, this will be explicitly stated.

In discussing positive definite forms, we shall first restrict our attention to the real case, though analogous results hold quite generally for hermitian forms.

§ 3. A Necessary Condition for Positive Definiteness

Let us begin by proving that if A is positive definite then $|A| > 0$. It is first of all easily seen that $|A| \neq 0$ if A is positive definite. Thus, if we should have $|A| = 0$, we could choose a nontrivial set of values x_j so that the equations

$$\sum_{j=1}^{n} a_{ij} x_j = 0 , \quad i = 1, 2, \ldots, n , \tag{1}$$

would be satisfied. For these values, we would have

$$(x, A\,x) = \sum_{i=1}^{n} x_i \left(\sum_{j=1}^{n} a_{ij} x_j \right) = 0 , \tag{2}$$

a contradiction.

To show that $|A| > 0$, we employ a continuity argument that can often be used in similar situations. Consider the matrix $\lambda I + (1 - \lambda) A$, where $0 \leqq \lambda \leqq 1$ and I is the identity matrix. Clearly this matrix is positive definite if A is positive definite. Hence $|\lambda I + (1 - \lambda) A|$ is nonzero. Since this determinant is continuous as a function of λ and positive at $\lambda = 1$, it necessarily is positive at $\lambda = 0$.

It follows that if A is positive definite, then all the determinants

$$|A_k| = |a_{ij}| , \ i, j = 1, 2, \ldots, k ; \ k = 1, 2, \ldots, n , \tag{3}$$

must be positive.

If it is given only that A is nonnegative definite, then $\lambda I + (1 - \lambda) A$ is positive definite for $0 < \lambda \leqq 1$, and the foregoing proof applies to show that the determinants (3) must all be nonnegative.

§ 4. Representation as a Sum of Squares

Let us now see if a positive definite quadratic form can be represented in a fashion that makes its positivity obvious. We shall see, in fact, that such a form can be exhibited as a sum of squares of linear functions. This type of investigation is in line with some of the remarks we have made in Chapter 1; see §§ 1.3 and 1.10. Obvious representations of this type are not always possible when we are dealing with general classes of functions; see [2.1], where further references will be found.

For $n = 2$, we may write the identity

$$a_{11} x_1^2 + 2 a_{12} x_1 x_2 + a_{22} x_2^2 = a_{11} \left(x_1 + \frac{a_{12} x_2}{a_{11}} \right)^2 + \left(a_{22} - \frac{a_{12}^2}{a_{11}} \right) x_2^2 , \tag{1}$$

provided that $a_{11} \neq 0$, whether or not the form is definite.

Note that

$$a_{22} - \frac{a_{12}^2}{a_{11}} = \frac{\begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix}}{a_{11}} . \tag{2}$$

The relationship (1) admits the following important extension:

Theorem 1. (Lagrange, Beltrami.) *If none of the determinants*

$$|A_k| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & & & \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{vmatrix}, \tag{3}$$

$k = 1, 2, \ldots, n-1$, *are zero, then* $Q(x) \equiv (x, Ax)$ *can be expressed in the form*

$$Q(x) = \sum_{k=1}^{n} \frac{|A_k|}{|A_{k-1}|} y_k^2, \tag{4}$$

where $|A_0| = 1$,

$$y_k = x_k + \sum_{j=k+1}^{n} b_{kj} x_j, \quad k = 1, 2, \ldots, n, \tag{5}$$

and the b_{kj} *are rational functions of the* a_{ij}.

This result is readily established via an inductive argument; see [2.1].

§ 5. A Necessary and Sufficient Condition for Positive Definiteness

As an immediate consequence of the representation (4) of § 4 and the results of § 3, we derive the following fundamental characterization:

Theorem 2. *A necessary and sufficient condition that the symmetric matrix* A *be positive definite is that*

$$|A_k| > 0, \quad k = 1, 2, \ldots, n. \tag{1}$$

Actually, for the necessity we can omit the assumption that A is symmetric since, as Ostrowski and Taussky [1] have shown, if $A = B + C$, where B is positive definite and C is skew symmetric, then $|A| \geqq |B|$.

The foregoing proof can readily be adapted to show that A is nonnegative definite if and only if the determinants in (1) are all nonnegative.

Theorem 2 plays an important role in the study of moment spaces. The method that is universally employed is that of reducing the positivity of a function to the positive definite character of a matrix. In this way, the relations (1) yield a set of necessary and sufficient conditions.

§ 6. Gramians

In the results of Theorem 2, we have a systematic method for generating chains of inequalities by means of matrices that obviously are positive definite. Consider, for example, a real, square matrix X, which we assume to be nonsingular; i. e., we assume that

$$\det X = |X| \neq 0.$$

If X' denotes the transpose of X, then XX' is symmetric and positive definite. This follows from the identity

$$(x, XX'x) = (X'x, X'x)\,, \tag{1}$$

and from the fact that, since X' is nonsingular, we have $X'x = 0$ only for the trivial vector $x = 0$.

Let X have the form

$$X = \begin{pmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(n)} \end{pmatrix}, \tag{2}$$

where the $x^{(i)}$ are row vectors. Then we may write

$$XX' = ((x^{(i)}, x^{(j)}))\,, \quad i, j = 1, 2, \ldots, n\,. \tag{3}$$

Accordingly, we see that if the vectors $x^{(i)}$ are linearly independent, then all the determinants, the *Gramians*,

$$G_k = |(x^{(i)}, x^{(j)})|\,, \quad i, j = 1, 2, \ldots, k\,, \tag{4}$$

$k = 1, 2, \ldots, n$, are positive.

The case $k = 2$ yields the result

$$\begin{vmatrix} \sum\limits_{i=1}^{n} x_i^2 & \sum\limits_{i=1}^{n} x_i y_i \\ \sum\limits_{i=1}^{n} x_i y_i & \sum\limits_{i=1}^{n} y_i^2 \end{vmatrix} > 0, \tag{5}$$

or

$$\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right) > \left(\sum_{i=1}^{n} x_i y_i\right)^2, \tag{6}$$

provided that the vectors x and y are linearly independent. This is the Cauchy inequality; see § 2 of Chapter 1.

Just as in the case of the Cauchy inequality, for which we have previously noted the Lagrange identity

$$\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right) - \left(\sum_{i=1}^{n} x_i y_i\right)^2 = \sum_{i,j=1}^{n} (x_i y_j - x_j y_i)^2, \tag{7}$$

there is a representation of G_k, for arbitrary k, as a sum of squares; see page 16 of HARDY, LITTLEWOOD, PÓLYA ([1.1] in Chapter 1).

Actually, the representation $A = TLT'$, where T is a real, orthogonal matrix and L is a diagonal matrix with positive elements down the main diagonal, is valid for any positive definite matrix A. It shows that any matrix of this type may be considered, relative to a generalized inner product, to be a Gramian. This observation furnishes another proof of Theorem 2.

A number of interesting results can be derived from the following representation theorem:

$$\left|\int_a^b f_i(x)\, g_j(x)\, dx\right| = \frac{1}{n!}\int_a^b \cdots \int_a^b \begin{vmatrix} f_1(x_1) & f_1(x_2) & \dots & f_1(x_n) \\ f_2(x_1) & f_2(x_2) & \dots & f_2(x_n) \\ \vdots \\ f_n(x_1) & f_n(x_2) & \dots & f_n(x_n) \end{vmatrix} \cdot \begin{vmatrix} g_1(x_1) & g_1(x_2) & \dots & g_1(x_n) \\ g_2(x_1) & g_2(x_2) & \dots & g_2(x_n) \\ \vdots \\ g_n(x_1) & g_n(x_2) & \dots & g_n(x_n) \end{vmatrix} dx\,, \tag{8}$$

where dx denotes $dx_1\, dx_2 \dots dx_n$; see PÓLYA and SZEGÖ ([44.1] in Chap. 1), page 48. See also ANDREIEF [1], DE BRUIJN [2], JACOBSON [3], MACDUFFEE [4], and KOLMOGOROFF [5]. From (8), it is clear that if we set

$$[f, g] = \left|\int_a^b f_i(x)\, g_j(x)\, dx\right|, \quad i, j = 1, 2, \dots, n\,, \tag{9}$$

then all the results valid for the usual scalar inner product (f, g) are valid for this generalized inner product. For example, an inequality due to DAVIS, given also by EVERITT [6], namely

$$[f, g]^2 \leqq [f, f]\,[g, g]\,, \tag{10}$$

is an immediate consequence. Further results may be found in MOFFERT [7].

Generally, we shall avoid any detailed discussion of inequalities involving determinants and minors. The interested reader may consult DE BRUIJN [8].

§ 7. Evaluation of an Infinite Integral

A number of the results that follow depend on the explicit evaluation of the multidimensional integral

$$J_n = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-(x, Ax)}\, dx\,, \tag{1}$$

where A is a positive definite matrix and the integration is over the entire real n-dimensional space.

Theorem 3. *If A is a positive definite matrix of order n, then the integral J_n can be expressed as*

$$J_n = \frac{\left(\int_{-\infty}^{\infty} e^{-x^2}\, dx\right)^n}{|A|^{1/2}} = \frac{\pi^{n/2}}{|A|^{1/2}}\,. \tag{2}$$

Proof. Using the representation of $Q(x) = (x, Ax)$ as a sum of squares as given in (4.4), perform the change of variables

$$y_k = x_k + \sum_{j=k+1}^{n} b_{kj} x_j\,, \quad k = 1, 2, \dots, n\,, \tag{3}$$

in the multiple integral. Since the value of the JACOBIAN of the transformation is 1, we obtain

$$J_n = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\sum_{j=1}^{n} (|A_k|/|A_{k-1}|)\, y_k^2}\, dy$$
$$= \frac{\left(\int_{-\infty}^{\infty} e^{-y^2} dy\right)^n}{\prod_{k=1}^{n} (|A_k|/|A_{k-1}|)^{1/2}} = \frac{\pi^{n/2}}{|A|^{1/2}}. \tag{4}$$

Actually, the precise value of the constant

$$\int_{-\infty}^{\infty} e^{-y^2} dy$$

is unimportant to us insofar as our subsequent use of this evaluation of J_n is concerned.

§ 8. Complex Matrices with Positive Definite Real Part

With a bit more effort, we can establish the following result:

Theorem 4. *If A and B are real, symmetric matrices of order n, with A positive definite, then*

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-(x, (A + iB)x)}\, dx = \frac{\pi^{n/2}}{|A + iB|^{1/2}}, \tag{1}$$

in which the principal values of the square roots are understood to be used.

An immediate consequence of this, foreshadowing our subsequent use of the identity (1), is the following result due to OSTROWSKI and TAUSSKY [6.1]; see also TAUSSKY [1] and BELLMAN and HOFFMAN [2].

Theorem 5. *If A and B are real, symmetric matrices with A positive definite, then the absolute value $||A + iB||$ of the determinant $|A + iB|$ satisfies the inequality*

$$||A + iB|| \geq |A|, \tag{2}$$

the sign of equality holding if and only if B is identically zero.

Proof. We have

$$\left| \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-(x, (A + iB)x)}\, dx \right| \leqq \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-(x, Ax)}\, dx, \tag{3}$$

in which the vertical bars denote absolute value. From (3), the result of Theorem 5 follows immediately.

§ 9. A Concavity Theorem

Again using the representation of Theorem 3, along the lines sketched in § 19 of Chapter 1, let us demonstrate a result of KY FAN [1].

Theorem 6. *If A and B are real, positive definite matrices, then*

$$|\lambda A + (1-\lambda) B| \geqq |A|^{\lambda} |B|^{1-\lambda} \tag{1}$$

for $0 \leqq \lambda \leqq 1$.

Proof. We have

$$\frac{\pi^{n/2}}{|\lambda A + (1-\lambda) B|^{1/2}} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\lambda (x, Ax) - (1-\lambda)(x, Bx)}\, dx . \tag{2}$$

Since the result is obviously valid for $\lambda = 0$ and for $\lambda = 1$, it is sufficient to consider values λ satisfying $0 < \lambda < 1$ Let us apply HÖLDER's inequality with exponents $p = 1/\lambda$, $q = 1/(1-\lambda)$ to the integral in (2). The resulting inequality is

$$\begin{aligned} &\frac{\pi^{n/2}}{|\lambda A + (1-\lambda) B|^{1/2}} \\ &\leqq \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-(x, Ax)}\, dx\right)^{\lambda} \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-(x, Bx)}\, dx\right)^{1-\lambda} \\ &\leqq \frac{\pi^{n/2}}{|A|^{\lambda/2} |B|^{(1-\lambda)/2}} , \end{aligned} \tag{3}$$

which is equivalent to (1). This proof is given in [2.1].

§ 10. An Inequality Concerning Minors

Continuing in the same vein, let us demonstrate the following result:

Theorem 7. *If A is a real, positive definite matrix of order n, then*

$$|A_{1n}| \leqq |A_{1k}| \cdot |A_{k+1,\, n}| , \tag{1}$$

where the determinant $|A_{rs}|$ is defined by

$$|A_{rs}| = |a_{ij}| , \quad i, j = r, r+1, \ldots, s . \tag{2}$$

In particular,

$$|A| \leqq a_{11} a_{22} \ldots a_{nn} . \tag{3}$$

Proof. In the integral for J_n, make the change of variables

$$\begin{aligned} x_i &= -x_i, \quad i = 1, 2, \ldots, k , \\ x_i &= x_i , \quad i = k+1, \ldots, n , \end{aligned} \tag{4}$$

and add the integral thus obtained to J_n.

The result is

$$2 J_n = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\sum\limits_{i,j=1}^{k} a_{ij} x_i x_j - \sum\limits_{i,j=k+1}^{n} a_{ji} x_i x_j} (U + U^{-1})\, dx , \tag{5}$$

where U is the exponential

$$U = e^{-\left(\sum_{i=1}^{k} \sum_{j=k+1}^{n} a_{ij} x_i x_j + \sum_{i=k+1}^{n} \sum_{j=1}^{k} a_{ij} x_i x_j\right)}. \tag{6}$$

Since for all positive U we have

$$U + U^{-1} \geqq 2, \tag{7}$$

from (5) we obtain the inequality

$$\begin{aligned} J_n \geqq & \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\sum_{i,j=1}^{k} a_{ij} x_i x_j} dx\right) \\ & \cdot \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\sum_{i,j=k+1}^{n} a_{ij} x_i x_j} dx\right), \end{aligned} \tag{8}$$

which yields (1).

Using (1) for $k = 1$, and proceeding inductively, we derive (3). For an alternative proof of (3), extended to positive definite hermitian matrices, see BECKENBACH [1]; see also § 14, below.

§ 11. Hadamard's Inequality

The inequality (10.3) permits us to obtain the most famous of all determinantal inequalities:

Theorem 8. *If* $|x_{ij}|$ *is a real determinant of order* n, *then*

$$||x_{ij}|| \leq \prod_{i=1}^{n} \left(\sum_{j=1}^{n} x_{ij}^2\right)^{1/2}. \tag{1}$$

Proof. Let $X = (x_{ij})$. Since the result is trivial if $|X| = 0$, let us assume that $|X| \neq 0$. Then XX' is positive definite. Applying (10.3), we obtain HADAMARD's *inequality* (1). See HADAMARD [1], [2]. The proof given here follows BELLMAN [3]. For alternative proofs, see MARCUS [4] and OSTROWSKI [5].

§ 12. Szasz's Inequality

HADAMARD's inequality, like the inequality connecting the arithmetic mean and the geometric mean, caught the fancy of mathematicians, with the result that there are a wide range of different proofs and numerous extensions of this result. See FISCHER [1], WILLIAMSON [2], SCHUR [3], and BUSH and OLKIN [4].

A particularly interesting extension due to SZÁSZ is the following: Let P_k denote the product of all principal k-rowed minors of A. If A is positive definite, then

$$P_1 \geqq P_2^{1/\binom{n-1}{1}} \geqq P_3^{1/\binom{n-1}{2}} \geqq \cdots \geqq P_{n-1}^{1/\binom{n-1}{n-2}} \geqq P_n \,. \tag{1}$$

A recent proof of (1) is due to MIRSKY [5].

§ 13. A Representation Theorem for the Determinant of a Hermitian Matrix

Let us now establish an analogue of Theorem 3.

Theorem 9. *If H is a positive definite hermitian matrix, then*

$$J_n(H) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(\bar{z}, Hz)}\, dx\, dy = \pi^n/|H| \,, \tag{1}$$

where $z = x + iy$, and dx and dy denote integration over the real n-dimensional volume.

Proof. The proof is in two stages. We begin by showing that

$$J_n(H) = \frac{\pi^n}{|A|\,|I + A^{-1} B A^{-1} B|^{1/2}} \,, \tag{2}$$

where $H = A + iB$, and A and B are real. Since

$$(\bar{z}, Hz) = (x, Ax) + 2\,(Bx, y) + (y, Ay) \,. \tag{3}$$

The result (2) is a consequence of a double application of Theorem 3; see page 61.

To complete the proof, observe that

$$\begin{aligned} |H| &= |A + iB| = |A|\,|I + iA^{-1}B| \,, \\ |H'| &= |A - iB| = |A|\,|I - iA^{-1}B| \,, \end{aligned} \tag{4}$$

and thus that

$$\begin{aligned} |H|\,|H'| = |H|^2 &= |A|^2\,|(I + iA^{-1}B)\,(I - iA^{-1}B)| \\ &= |A|^2\,|I + A^{-1}BA^{-1}B| \,. \end{aligned} \tag{5}$$

This result is given by BELLMAN [1], partially extending a theorem of HUA [2].

§ 14. Discussion

Having obtained Theorem 9, we can now proceed, as in §§ 7—12, to derive extensions, to hermitian matrices, of the preceding results stated only for real, symmetric matrices.

In the following section, we shall discuss some more versatile representations due to INGHAM, SIEGEL, BELLMAN, and OLKIN.

§ 15. Ingham-Siegel Integrals and Generalizations

We have seen in the foregoing sections that a number of interesting results can be obtained from the quasi-linear representation for $|A|^{-1/2}$.

It is natural then to ask whether or not further results of this type can be obtained from other representations.

To obtain these representations, we turn to integrals introduced by SIEGEL [2] in connection with his theory of matrix modular functions, and by INGHAM [1] in connection with statistical problems of multivariate analysis.

The classical integral of EULER is

$$\int_0^\infty e^{-xy} x^{s-1}\, dx = \Gamma(s)\, y^{-s}, \ \Re(s), \Re(y) > 0\,. \tag{1}$$

An extensive generalization of this, due to SIEGEL [2], is

$$\begin{aligned} &\int_{X>0} e^{-\mathrm{tr}\,(XY)}\, |X|^{\,s-(n+1)/2}\, dV \\ &= \pi^{n\,(n-1)/4} \frac{\Gamma(s)\,\Gamma(s-1/2)\cdots\Gamma(s-(n-1)/2)}{|Y|^s}\,. \end{aligned} \tag{2}$$

Here X and Y are positive definite matrices of order n, $\mathrm{tr}(X)$ denotes the trace of X,

$$dV = \prod_{i \leqq j} dx_{ij}\,,$$

and the integration is over the region where X is positive definite. The real part of s is to be greater than $(n-1)/2$.

The integral equivalent to the Laplace inverse of (2) was given by INGHAM [1] in connection with a problem arising in multivariate analysis.

The integrals of INGHAM and SIEGEL, in turn, can be extended, following a suggestion of SELBERG; see BELLMAN [6] and OLKIN [3]. A typical result is

$$\begin{aligned} &\int_{X>0} \frac{|X|^{\sum_{i=1}^{n} k_i-(n+1)/2}\; e^{-\mathrm{tr}\,(XY)}}{|X^{(2)}|^{k_1}\,|X^{(3)}|^{k_2}\cdots|X^{(p)}|^{k_{n-1}}}\, dV \\ &= \frac{(\sqrt{\pi})^{n(n-1)/2}\,\Gamma(k_n)\,\Gamma(k_n+k_{n-1}-1/2)\cdots\Gamma\left[\sum_{i=1}^{n} k_i-\frac{n}{2}+\frac{1}{2}\right]}{|Y_n|^{k_n}\,|Y_{n-1}|^{k_{n-1}}\cdots|Y_1|^{k_1}}\,. \end{aligned} \tag{3}$$

Here

$$\begin{aligned} |X^{(k)}| &= |x_{ij}|\,, \quad i,j = k, k+1, \ldots, n\,, \\ |Y_k| &= |y_{ij}|\,, \quad i,j = 1, 2, \ldots, k\,. \end{aligned} \tag{4}$$

The restriction on the k_i is that each of the expressions $k_n + k_{n-1} + \cdots + k_r - r/2$ be positive.

A number of further extensions were given by OLKIN [7]; see also BOCHNER ([39.3] in Chapter 1), GARDING [4], and AITKIN [5].

We shall not prove any of these identities, since the results will not be used. It is clear that we can establish a variety of concavity theorems using the representation (3). It turns out, however, that more extensive results can readily be established by use of a simpler technique; see BELLMAN [6].

Finally, let us mention the representation used by HUA to obtain additional determinantal inequalities. Some of his results follow also from the representation of § 13; see BELLMAN [13.1].

§ 16. Group Invariance and Representation Formulas

In his paper [39.3] in Chapter 1, BOCHNER discussed group invariance and a number of representation formulas of the foregoing general type, demonstrating their common origin. Thus, for example, he derived the formula

$$\begin{aligned} &\int_R e^{-(x_1 y_1 + x_2 y_2 + \cdots + x_n y_n)}\, d x_1\, d x_2 \cdots d x_n \\ &= c_n\, (y_1^2 - y_2^2 - \cdots - y_n^2)^{-n/2}, \end{aligned} \tag{1}$$

where R is the region in (x, y) space defined by

$$x_1 > (x_2^2 + \cdots + x_n^2)^{1/2}, \quad y_1 > (y_2^2 + \cdots + y_n^2)^{1/2}, \tag{2}$$

$$c_n = 2\pi^{(n+1)/2}\, \Gamma(n-1)/\Gamma[(n+1)/2]\,. \tag{3}$$

From this, we can readily demonstrate the result given in § 39 of Chapter 1.

§ 17. Bergstrom's Inequality

As another example of the quasi-linearization technique, let us establish the following result of BERGSTROM [1]; the proof is due to BELLMAN [2].

Theorem 10. *Let A and B be positive definite matrices, and let A_i, B_i denote the submatrices obtained by deleting the i-th row and column. Then*

$$\frac{|A+B|}{|A_i + B_i|} \geqq \frac{|A|}{|A_i|} + \frac{|B|}{|B_i|}\,. \tag{1}$$

Proof. Consider the problem of minimizing the positive definite quadratic form

$$Q(x) = \sum_{i,j=1}^{n} a_{ij} x_i x_j = (x, A x)\,, \tag{2}$$

over all x_k, $k = 1, 2, \ldots, n$, subject to the restriction that $x_i = 1$.

It is easy to see, via a Lagrange multiplier or otherwise, that

$$\min_{x_i = 1} Q(x) = \frac{|A|}{|A_i|}\,. \tag{3}$$

We thus have a quasi-linear representation for $|A|/|A_i|$, whence the inequality (1) follows immediately.

§ 18. A Generalization

Let us now demonstrate a simultaneous generalization, suggested by KY FAN, of the foregoing result and of Theorem 6 of § 9.

Let $A^{(j)}$ denote the principal submatrix of A obtained by deleting the first $(j-1)$ rows and columns of A; in particular, let $A^{(1)} = A$. Let $B^{(j)}$ and $C^{(j)}$ have similar meanings, with $C = \lambda A + (1 - \lambda) B$, $0 \leqq \lambda \leqq 1$. It is assumed that A and B are positive definite. Then we have the following result.

Theorem 11. *Under the foregoing conditions, the inequality*

$$\prod_{j=1}^{n} |C^{(j)}|^{k_j} \geqq \prod_{j=1}^{n} |A^{(j)}|^{\lambda k_j} \, |B^{(j)}|^{(1-\lambda) k_j} \tag{1}$$

holds for any set of n real numbers k_i such that

$$\sum_{i=1}^{j} k_i \geqq 0 \,, \quad j = 1, 2, \ldots, n \,. \tag{2}$$

Proof. According to BERGSTROM's inequality in § 17, we have

$$\begin{aligned} \frac{|C^{(j)}|}{|C^{(j+1)}|} &\geqq \frac{|\lambda A^{(j)}|}{|\lambda A^{(j+1)}|} + \frac{|(1-\lambda) B^{(j)}|}{|(1-\lambda) B^{(j+1)}|} \\ &\geqq \lambda \frac{|A^{(j)}|}{|A^{(j+1)}|} + (1-\lambda) \frac{|B^{(j)}|}{|B^{(j+1)}|} \\ &\geqq \left(\frac{|A^{(j)}|}{|A^{(j+1)}|} \right)^{\lambda} \left(\frac{|B^{(j)}|}{|B^{(j+1)}|} \right)^{1-\lambda}, \end{aligned} \tag{3}$$

the last inequality holding by the arithmetic-mean — geometric-mean inequality.

The stated inequality (1) follows when we write

$$\begin{aligned} \prod_{j=1}^{n} |C^{(j)}|^{k_j} = {} & \left(\frac{|C^{(1)}|}{|C^{( )}|} \right)^{k_1} \left(\frac{|C^{(2)}|}{|C^{(3)}|} \right)^{k_1 + k_2} \cdots \\ & \left(\frac{|C^{n-1}|}{|C^{(n)}|} \right)^{k_1 + k_2 + \cdots + k_{n-1}} |C^{(n)}|^{k_1 + k_2 + \cdots + k_n} \end{aligned} \tag{4}$$

and use the inequality (3).

§ 19. Canonical Form

So far, we have been concentrating on the determinant of the matrix A and on its various minors. We now wish to study the characteristic roots of A as functions of A. In order to do this, we require the fundamental connection between A and its characteristic roots, which we write as $\lambda_1, \lambda_2, \ldots, \lambda_n$. If A is positive definite, we know that all the λ_i are positive. Let us then order them so that

$$\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_n \,. \tag{1}$$

There exists an orthogonal matrix T such that

$$T' A T = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & 0 \\ & 0 & \ddots & \\ & & & \lambda_n \end{pmatrix}. \tag{2}$$

Equivalently, there is an orthogonal transformation, $x = Ty$, such that

$$(x, A x) = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 , \tag{3}$$

another representation as a sum of squares if A is positive definite.

Similarly, if H is hermitian, we can find a unitary matrix U such that

$$U^* H U = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & 0 \\ & 0 & \ddots & \\ & & & \lambda_n \end{pmatrix}, \tag{4}$$

where U^* is the transpose of the matrix conjugate to U. Proofs of these results may be found in [**2.1**].

The identity (3) shows that a necessary and sufficient condition that a real symmetric matrix A be positive definite is that its characteristic roots be postive.

§ 20. A Generalization of Bergstrom's Inequality

Using the fundamental canonical form (2) of § 19, we can derive the following inequality, of which the Bergstrom inequality is a special case.

Theorem 12. *If the real matrix A is positive definite, then*

$$(x, A x)\,(y, A^{-1} y) \geqq (x, y)^2 \tag{1}$$

for all real x and y.

Proof. Reducing A to diagonal form by means of the orthogonal transformation $x = Tu$, and setting $y = Tv$, we have to show that

$$\left(\sum_{i=1}^{n} \lambda_i u_i^2\right)\left(\sum_{i=1}^{n} \frac{v_i^2}{\lambda_i}\right) \geqq (Tu, Tv)^2 = (u, v)^2. \tag{2}$$

This result, however, is a special case of the Cauchy inequality.

Since equality is obtained for a suitable choice of x, we can write

$$\psi(A) = (y, A^{-1} y)^{-1} = \min_x \frac{(x, A x)}{(x, y)^2} . \tag{3}$$

From this it is clear that

$$\psi(A + B) \geqq \psi(A) + \psi(B) . \tag{4}$$

If y is chosen to be the vector with the components

$$y_i = 1,\ y_j = 0,\ j \neq i , \tag{5}$$

we obtain Theorem 10.

As in Chapter 1, we can refine Theorem 12 in various ways; see GREUB and RHEINBOLDT [44.4] in Chapter 1. Thus, for example, we have the following result.

Theorem 13. *Let A and B be real, positive definite matrices of order n. If $m_1 I \leqq A \leqq m_2 I$, and $m_1, m_2 > 0$, then*

$$(x, x)^2 \leqq (A x, x)(A^{-1} x, x) \leqq \frac{(m_1 + m_2)^2}{4 m_1 m_2} (x, x)^2. \tag{6}$$

If $AB = BA$, $m_1 I \leqq A \leqq m_2 I$, $m_3, m_4 > 0$, and $m_3 I \leqq B \leqq m_4 I$, then

$$(A x, A x)(B x, B x) \leqq \frac{(m_1 m_3 + m_2 m_4)^2}{m_1 m_2 m_3 m_4} (A x, B x)^2. \tag{7}$$

§ 21. A Representation Theorem for $|A|^{1/n}$

Prior to presenting an inequality of MINKOWSKI, which admits a number of interesting extensions, let us establish a representation theorem of a type different from any of the preceding ones.

Theorem 14. *If A is a real, positive definite matrix of order n, then*

$$|A|^{1/n} = \min_{|B|=1} \operatorname{tr}(A B)/n, \tag{1}$$

where B is positive definite.

Proof. In view of the identity $\operatorname{tr}(T' A T B) = \operatorname{tr}(A T B T')$, it is sufficient to consider that A is in diagonal form. Then

$$\operatorname{tr}(A B) = \sum_{i=1}^{n} \lambda_i b_{ii}. \tag{2}$$

Using the arithmetic-mean — geometric-mean inequality, we see that

$$\frac{\operatorname{tr}(A B)}{n} \geqq \left(\prod_{i=1}^{n} \lambda_i\right)^{1/n} \left(\prod_{i=1}^{n} b_{ii}\right)^{1/n} = |A|^{1/n} \left(\prod_{i=1}^{n} b_{ii}\right)^{1/n}. \tag{3}$$

Referring to § 10, we see that

$$\prod_{i=1}^{n} b_{ii} \geqq |B| = 1,$$

whence (1) follows.

§ 22. An Inequality of Minkowski

From the representation (21.1), we obtain immediately a result due to MINKOWSKI:

Theorem 15. *If A and B are positive definite matrices of order n, then*

$$|A + B|^{1/n} \geqq |A|^{1/n} + |B|^{1/n}. \tag{1}$$

For an extension of this result to matrices that are not necessarily symmetric, see HAYNESWORTH [1], [2], where many additional reference will be found.

§ 23. A Generalization due to Ky Fan

KY FAN [1] has given a simultaneous generalization of the foregoing inequality (22.1), and of the inequality of BERGSTROM referred to above in § 17, namely the following:

Theorem 16. *Let* A_k *denote the principal submatrix of* A *formed by taking the first* k *rows and columns of* A. *If* $C = A + B$, *where* A *and* B *are positive definite matrices of order* n, *then*

$$\left(\frac{|C|}{|C_k|}\right)^{1/(n-k)} \geqq \left(\frac{|A|}{|A_k|}\right)^{1/(n-k)} + \left(\frac{|B|}{|B_k|}\right)^{1/(n-k)}. \tag{1}$$

The proof is based on a minimum property of a type that generalizes (17.3). See also KY FAN [2] and MIRSKY [3].

§ 24. A Generalization due to Oppenheim

A generalization of a different type is due to OPPENHEIM [2]; see also his earlier paper [1].

Theorem 17. *Let* A *and* B *be positive definite matrices of order* n, *and let* $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_n$, $\mu_1 \geqq \mu_2 \geqq \cdots \geqq \mu_n$, $\nu_1 \geqq \nu_2 \geqq \cdots \geqq \nu_n$ *be the characteristic roots of* A, B, *and* $A + B$, *respectively. Then*

$$\left(\prod_{i=1}^{k} \nu_i\right)^{1/k} \geqq \left(\prod_{i=1}^{k} \lambda_i\right)^{1/k} + \left(\prod_{i=1}^{k} \mu_i\right)^{1/k}, \tag{1}$$

for $k = 1, 2, \ldots, n$.

This result may be derived either by reduction to diagonal form or by a representation of the type appearing in Theorem 25 of § 34, below.

§ 25. The Rayleigh Quotient

The characteristic roots are initially determined as the roots of the polynomial equation

$$|A - \lambda I| = 0,$$

the *characteristic equation* of A. If A is real and symmetric, we know that the λ_i are all real.

Since it is not at all easy to obtain the properties of the λ_i as functions of A from this description, we look about for other characterizations. The key to these is the representation in diagonal form given in § 19, together with a variational characterization of the λ_i. From either the physical point of view (characteristic frequencies) or the geometric picture (axes of an ellipsoid), we are led to introduce the Rayleigh quotient, $(x, A\,x)/(x, x)$.

We can then state a striking result:

Theorem 18. *If A is a real, symmetric matrix, then*

$$\begin{aligned} \lambda_1 &= \max_x \frac{(x, A\,x)}{(x, x)}, \\ \lambda_n &= \min_x \frac{(x, A\,x)}{(x, x)}. \end{aligned} \tag{1}$$

The proof can be obtained in many ways. One method uses the representation of § 19, namely,

$$(x, A\,x) = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2,$$

for $(x, x) = (y, y)$, $x = T\,y$.

When λ_1 and λ_n are regarded as functions of A, the foregoing quasi-linear representation (1) shows that if B is a nonnegative definite matrix, then

$$\begin{aligned} \lambda_1\,(A + B) &\geqq \lambda_1\,(A)\,, \\ \lambda_n\,(A + B) &\geqq \lambda_n(A)\,. \end{aligned} \tag{2}$$

Furthermore, (1) yields the inequalities

$$\begin{aligned} \lambda_1\,(\lambda A + (1-\lambda)\,B) &\leqq \lambda\lambda_1(A) + (1-\lambda)\,\lambda_1(B)\,, \\ \lambda_n\,(\lambda A + (1-\lambda)\,B) &\geqq \lambda\lambda_n(A) + (1-\lambda)\,\lambda_n(B)\,, \end{aligned} \tag{3}$$

for any two symmetric matrices A and B. In other words, $\lambda_1(A)$ is a convex function of A, while $\lambda_n(A)$ is a concave function of A.

§ 26. The Fischer Min-max Theorem

The representation given in Theorem 18 can be extended in the following fashion. Let $x^{(1)}$ be a characteristic vector associated with λ_1, normalized by the condition that $(x^{(1)}, x^{(1)}) = 1$. Then, from either analytic or geometric considerations, it is clear that

$$\lambda_2 = \max_x \frac{(x, A\,x)}{(x, x)}, \tag{1}$$

where x is constrained by the orthogonality condition $(x, x^{(1)}) = 0$. Let $x^{(2)}$ be a normalized characteristic vector associated with λ_2, yielding the maximum in (1). Then we can write

$$\lambda_3 = \max_x \frac{(x, A\,x)}{(x, x)}, \tag{2}$$

where x is now constrained by the two orthogonality conditions, $(x, x^{(1)}) = 0$, $(x, x^{(2)}) = 0$, and so on.

The difficulty in using this apparently quasi-linear representation of the λ_i lies in the fact that the characteristic vectors, $x^{(1)}, x^{(2)}, \ldots$, also depend on A. Consequently, in place of this inductive definition of the λ_i, we need a representation that yields λ_i without implicit or explicit dependence on $\lambda_1, \lambda_2, \ldots, \lambda_{i-1}$.

Such a representation was obtained by FISCHER; it plays a vital role in the further analytic theory of matrices. The generalization, yielding corresponding results for the characteristic values of wide classes of symmetric operators, is due to COURANT [1]; for FISCHER's original paper, see [2].

Theorem 19. *For any real, symmetric matrix A, the characteristic roots $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_n$ may be obtained as follows:*

$$\begin{aligned} \lambda_1 &= \max_{(x,x)=1} (x, A x) , \\ \lambda_2 &= \min_{(y,y)=1} \max_{\substack{(x,x)=1 \\ (x,y)=0}} (x, A x) , \\ &\vdots \\ \lambda_k &= \min_{\substack{(y^{(i)}, y^{(i)})=1 \\ i=1,2,\ldots,k-1}} \max_{\substack{(x,x)=1 \\ (x,y^{(i)})=0 \\ i=1,2,\ldots,k-1}} (x, A x) . \end{aligned} \tag{1}$$

Equivalently, the roots may be represented as

$$\begin{aligned} \lambda_n &= \min_{(x,x)=1} (x, A x) , \\ \lambda_{n-1} &= \max_{(y,y)=1} \min_{\substack{(x,x)=1 \\ (y,x)=0}} (x, A x) , \end{aligned} \tag{2}$$

and so on.

For proof of Theorem 19, we refer to [1] and to [2.1].

From this result, we readily see that

$$\lambda_k (A + B) \geqq \lambda_k(A) , \quad k = 1, 2, \ldots, n , \tag{3}$$

for any real, symmetric matrix A and any real, nonnegative definite matrix B, a monotonicity result that is physically obvious since "stiffening" a rod or plate increases all of its characteristic frequencies.

§ 27. A Representation Theorem

Let us introduce the notation

$$|A|_k = \lambda_n \lambda_{n-1} \cdots \lambda_{n-k+1} , \tag{1}$$

the product of the first k *smallest* characteristic roots of the real, positive definite matrix A.

Denote by R_k a k-dimensional subspace of the n-dimensional x_i space defined by the $n-k$ relations

$$(x, a_i) = 0 , \quad i = 1, 2, \ldots, n-k , \tag{2}$$

where the a_i are $n-k$ linearly independent vectors.

Then the result we wish to demonstrate is this

Theorem 20. *For any real, positive definite matrix A,*

$$\frac{\pi^{k/2}}{|A|_k^{1/2}} = \max_{R_k} \int_{R_k} e^{-(x, Ax)} d V_k , \tag{3}$$

where the integration is over a k-dimensional subspace defined by (2), *and the maximization is over all such R_k.*

Proof. It is clear that we can begin by taking $(x, A x)$ to have the form $\lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2$. Consider the volume, $V_a(\varrho)$, contained in the region determined by the condition

$$\begin{aligned} &\lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2 \leqq \varrho , \\ &(x, a_i) = 0 , \quad i = 1, 2, \ldots, n-k . \end{aligned} \tag{4}$$

Then, clearly,

$$V_a(\varrho) = \varrho^{k/2} V_a(1) . \tag{5}$$

Hence,

$$\begin{aligned} \int_{R_k} e^{-\sum_{i=1}^{n} \lambda_i x_i^2} d V_k &= \int_{-\infty}^{\infty} e^{-\varrho} d V_a(\varrho) \\ &= \frac{k}{2} V_a(1) \left(\int_{-\infty}^{\infty} e^{-\varrho} \varrho^{\frac{k}{2}-1} d\varrho \right). \end{aligned} \tag{6}$$

To complete the proof, we must show that the maximum of $V_a(1)$ is obtained when the relations (2) are

$$x_1 = x_2 = \cdots = x_{n-k} = 0 . \tag{7}$$

This, however, is an immediate consequence of the formula for the volume of the ellipsoid described by the quadratic form

$$(x, B x) = 1 , \tag{8}$$

where B is positive definite, and of the results of FISCHER given in § 26.

§ 28. An Inequality of Ky Fan

With the aid of Theorem 20 and the method employed in § 9, we establish the following concavity theorem of FAN [1].

Theorem 21. *If A and B are real, positive definite matrices of order n, and $0 \leqq \lambda \leqq 1$, then*

$$|\lambda A + (1-\lambda) B|_k \geqq |A|_k^{\lambda} |B|_k^{(1-\lambda)} \tag{1}$$

FAN's proof is based on the representation given in § 32, below.

§ 29. An Additive Version

Let us now obtain an additive version of Theorem 21. If in place of A we consider the matrix $I + \varepsilon A$, and in place of B the matrix $I + \varepsilon B$, for $\varepsilon > 0$, the result of Theorem 21 reads

$$|I + \varepsilon (\lambda A + (1-\lambda) B)|_k \geqq |I + \varepsilon A|_k^{\lambda} \, |I + \varepsilon B|_k^{1-\lambda}. \tag{1}$$

The characteristic roots of $I + \varepsilon A$ are $1 + \varepsilon\lambda_1 \geqq 1 + \varepsilon\lambda_2 \geqq \ldots \geqq 1 + \varepsilon\lambda_n$, and similarly those of $I + \varepsilon B$ are $1 + \varepsilon\mu_1 \geqq 1 + \varepsilon\mu_2 \geqq \ldots \geqq 1 + \varepsilon\mu_n$. Hence, for small positive ε we obtain the result that

$$\begin{aligned} |I + \varepsilon A|_k^{\lambda} &= 1 + \lambda\varepsilon\,(\lambda_n + \lambda_{n-1} + \cdots + \lambda_{n-k+1}) + O(\varepsilon^2)\,, \\ |I + \varepsilon B|_k^{1-\lambda} &= 1 + (1-\lambda)\,\varepsilon\,(\mu_n + \mu_{n-1} + \cdots + \mu_{n-k+1}) + O(\varepsilon^2)\,. \end{aligned} \tag{2}$$

Letting $\varepsilon \to 0$, from the inequality of (28.1) we obtain a further inequality:

Theorem 22. *Let*

$$S_k(A) = \lambda_n + \lambda_{n-1} + \cdots + \lambda_{n-k+1} \tag{3}$$

for any real, symmetric matrix A. If A and B are matrices of order n, and $0 \leqq \lambda \leqq 1$, then

$$S_k(\lambda A + (1-\lambda) B) \geqq \lambda S_k(A) + (1-\lambda)\, S_k(B)\,. \tag{4}$$

This result is also due to FAN [9.1]. The proof given here follows BELLMAN [2.1].

§ 30. Results Connecting Characteristic Roots of A, AA^*, and $(A + A^*)/2$

In addition to the results described in §§ 19—29, a great many relations have been established between the characteristic roots of A and those of AA^* and $(A + A^*)/2$. Here A^* is the conjugate of the transpose of A, $A^* = \overline{A'}$.

For an account of these results and additional sources, the reader is referred to the papers by WEYL [1], KY FAN [2], LIDSKII [3], HORN [4], WIELANDT [5], AMIR-MOEZ [6], and MIRSKY [7].

Observe that, as in the foregoing sections, every inequality for AA^* yields a corresponding inequality for $(A + A^*)/2$.

§ 31. The Cauchy-Poincaré Separation Theorem

An immediate consequence of FISCHER's min-max characterization (§ 26) is the following result of CAUCHY [1] and POINCARÉ [2], which we shall use in § 32.

Theorem 23. *Let* $\{x^{(j)}\}$, $j = 1, 2, \ldots, k \leqq n$, *be a set of k orthonormal vectors, and set*

$$x = \sum_{j=1}^{k} u_k x^{(j)} . \tag{1}$$

Then

$$(x, A x) = \sum_{j,l=1}^{k} u_j u_l (x^{(j)}, A x^{(l)}) . \tag{2}$$

Consider the k-dimensional symmetric matrix B defined by

$$B = ((x^{(j)}, A x^{(l)})) , \quad j, l = 1, 2, \ldots, k . \tag{3}$$

Then

$$\begin{aligned} \lambda_i(B) &\leqq \lambda_i(A) , \quad i = 1, 2, \ldots, k , \\ \lambda_{k-j}(B) &\geqq \lambda_{n-j}(A) , \quad j = 0, 1, 2, \ldots, k-1 . \end{aligned} \tag{4}$$

See also PÓLYA [3] and HAMBURGER and GRIMSHAW [4].

§ 32. An Inequality for $\lambda_n \lambda_{n-1} \ldots \lambda_k$

Using the foregoing theorem, we can demonstrate the following result of KY FAN [1]:

Theorem 24. *Let A be a positive definite matrix. Then*

$$\lambda_n \lambda_{n-1} \ldots \lambda_k = \min \prod_{i=k}^{n} (x^{(i)}, A x^{(i)}) , \quad k = 1, 2, \ldots, n , \tag{1}$$

where the minimum is taken over all $n - k + 1$ *orthonormal vectors* $\{x^{(i)}\}$.

Proof. We know from Theorem 7 of § 10 that

$$\prod_{i=1}^{k} \lambda_i(B) = |B| \leqq \prod_{t=1}^{k} (x^{(i)}, A x^{(i)}) . \tag{2}$$

Since $\lambda_k(B) \geqq \lambda_n(A)$, $\lambda_{k-1}(B) \geqq \lambda_{n-1}(A)$, and so on, as in (31.4), the inequality (2) yields the desired result, since equality can actually be obtained by suitable choice of the $x^{(i)}$. Reference to further results will be found in § 36.

§ 33. Discussion

It is clear that there are many interrelations between a number of the results we have obtained. For the benefit of the reader, let us present a "flow chart." The logical relations are

$$\text{Theorem 24} \begin{cases} \nearrow \text{Theorem 21} \longrightarrow \text{Theorem 6} \\ \searrow \text{Theorem 5} \end{cases}$$

and, as we shall see in § 34,

$$\text{Theorem 24} \longrightarrow \text{Theorem 25} \longrightarrow \text{Theorem 22.}$$

§ 34. Additive Version

Replacing A by $I + \varepsilon A$, and proceeding as in § 21, we readily derive the additive complement of Theorem 24, also due to FAN [9.1]; see MIRSKY [1].

Theorem 25. *For any real, symmetric matrix A of order n,*

$$\begin{aligned} \sum_{i=1}^{k} \lambda_i &= \max \sum_{i=1}^{k} (x^{(i)}, A\, x^{(i)}) , \\ \sum_{i=1}^{k} \lambda_{n-i+1} &= \min \sum_{i=1}^{k} (x^{(i)}, A\, x^{(i)}) , \end{aligned} \tag{1}$$

where in both cases the variation is over all sets of k orthonormal vectors.

Theorem 22 is a direct consequence of this representation.

Observe that it is not necessary to impose the condition that A be positive definite, since $I + \varepsilon A$ is positive definite for ε sufficiently small. An immediate consequence of Theorem 25 is the analogue of (10.3),

$$\sum_{i=k}^{n} \lambda_i \leq \sum_{i=k}^{n} a_{ii}, \quad k = 1, 2, \ldots, n . \tag{2}$$

§ 35. Multiplicative Inequality Derived from Additive

Let us now show that not only can we derive the additive inequality from the multiplicative, but that also we can derive the multiplicative one from the additive; see BELLMAN [1]. In what follows, the matrix A will be positive definite.

Consider the following summation by parts:

$$\begin{aligned} c_k \lambda_k + c_{k+1} \lambda_{k+1} + \cdots + c_n \lambda_n = {} & c_k (\lambda_k + \lambda_{k+1} + \cdots + \lambda_n) \\ & + (c_{k+1} - c_k) (\lambda_{k+1} + \cdots + \lambda_n) \\ & + \cdots + (c_n - c_{n-1}) \lambda_n . \end{aligned} \tag{1}$$

Assume for the moment that $0 \leq c_k \leq c_{k+1} \leq \cdots \leq c_n$. Then, referring to § 34, we see that

$$\begin{aligned} c_k \lambda_k + c_{k+1} \lambda_{k+1} + \cdots + c_n \lambda_n \leq {} & c_k \left[\sum_{i=k}^{n} (x^{(i)}, A\, x^{(i)}) \right] \\ & + (c_{k+1} - c_k) \left[\sum_{i=k+1}^{n} (x^{(i)}, A\, x^{(i)}) \right] \\ & + \cdots + (c_n - c_{n-1}) (x^{(i)}, A\, x^{(n)}) \\ \leq {} & c_k (x^{(k)}, A\, x^{(k)}) + c_{k+1} (x^{(k+1)}), A\, x^{(k+1)}) + \cdots + c_n (x^{(n)}, A\, x^{(n)}), \end{aligned} \tag{2}$$

for any set of orthonormal vectors $\{x^{(i)}\}$.

From (2) we deduce that

$$\min_R \left(\sum_{i=k}^{n} c_i \lambda_i \right) \leqq \min_R \left[\sum_{i=k}^{n} c_i \, (x^{(i)}, A\, x^{(i)}) \right], \tag{3}$$

where R is the region in c_i space defined by the relations

$$\begin{aligned} &\text{(a)} \quad 0 < c_k \leqq c_{k+1} \leqq \cdots \leqq c_n , \\ &\text{(b)} \quad \prod_{i=k}^{n} c_i = 1 . \end{aligned} \tag{4}$$

The arithmetic-mean — geometric-mean inequality, which we thoroughly proved in Chapter 1, tells us that

$$\frac{\sum_{i=k}^{n} c_i \lambda_i}{n-k+1} \geqq \left(\prod_{i=k}^{n} c_i \right)^{1/(n-k+1)} \left(\prod_{i=k}^{n} \lambda_i \right)^{1/(n-k+1)}, \tag{5}$$

and thus that the minimum of $\sum_{i=k}^{n} c_i \lambda_i$ is assumed at the points

$$c_i = \frac{\left(\sum_{i=k}^{n} \lambda_i \right)^{1/(n-k+1)}}{\lambda_i}, \quad i = k, \ldots, n . \tag{6}$$

Since $\lambda_k \geqq \lambda_{k+1} \geqq \cdots \geqq \lambda_n$, we see that the restriction $c_k \leqq c_{k+1} \leqq \cdots \leqq c_n$ does not exclude the minimum point on the left-hand side. This is a device we used in § 30 of Chapter 1. In order to maintain this restriction on the right-hand side, let us temporarily impose the additional restriction

$$(x^{(k)}, A\, x^{(k)}) \leqq (x^{(k+1)}, A\, x^{(k+1)}) \leqq \cdots \leqq (x^{(n)}, A\, x^{(n)}) . \tag{7}$$

Then (3) yields the relation

$$\prod_{i=k}^{n} \lambda_i \leqq \prod_{i=k}^{n} (x^{(i)}, A\, x^{(i)}) \tag{8}$$

for any orthonormal set $\{x^{(i)}\}$ satisfying (7).

From the symmetry of the condition, we see that this is no essential restriction.

§ 36. Further Results

The results of Ky Fan concerning variational characterizations of such functions as $\lambda_n \lambda_{n-1} \cdots \lambda_k$ and $\lambda_n + \lambda_{n-1} + \cdots + \lambda_k$ have been extensively generalized. In the first place, we can ask for a representation of a sum such as $\lambda_1 + \lambda_3 + \lambda_{10}$. In the second place, we can ask for a representation of a product such as $\lambda_1 \lambda_3 \lambda_{10}$.

Next, we can consider more general symmetric functions, such as $\sum_{i \neq j} \lambda_i \lambda_j$, and so on. For a number of results of this nature, see Ostrowski

[28.1], MARCUS and LOPES [33.1] (both in Chapter 1), MARCUS and MCGREGOR [1], MARCUS and MOYLS [2], MARCUS, MOYLS, and WESTWICK [3], [4], and ALI R. AMIR-MOEZ [7], where results of WIELANDT are extensively generalized.

§ 37. Compound and Adjugate Matrices

Associated with a linear transformation, $y = A\,x$, there are a number of important associated or induced transformations, which can often be effectively used to study various properties of the original transformation [1]. Of these, perhaps the most important are the adjoint transformations.

The adjoint operator, A', is obtained by means of the relation $(A\,x, y) = (x, A'y)$. This *defines* the matrix A' in the finite-dimensional case, and the same technique is used in more general situations.

Other important transformations, "induced transformations," are obtained by considering certain functions of x and studying the transformations effected upon the quantities by means of the relations $y = A\,x$.

For example, given the two transformations

$$y_i = \sum_{j=1}^{n} a_{ij} x_j, \; w_i = \sum_{i=1}^{n} b_{ij} z_j, \quad i = 1, 2, \ldots, n\,, \tag{1}$$

we may write

$$y_i w_j = \sum_{k,\,r=1}^{n} a_{ik} b_{jr} x_k z_r\,. \tag{2}$$

Hence, if we introduce the two n^2-dimensional vectors with components $y_i w_j$ and $x_i z_j$, $i, j = 1, 2, \ldots, n$, these are related by a matrix, which we call the *Kronecker product* of A and B, the n^2-dimensional matrix

$$A \times B = (a_{ij} B)\,. \tag{3}$$

See [2], [3], and [2.1] for further details and references.

In place of these simple functions, we can introduce higher-dimensional products. A particularly interesting and important choice is the set of 2 by 2 determinants

$$\begin{vmatrix} x_i & y_i \\ x_j & y_j \end{vmatrix}, \quad i, j = 1, 2, \ldots, n\,, \tag{4}$$

and generally the r by r determinants formed in an analogous fashion. These introduce the *compound* or *adjugate* matrices of I. SCHUR [1] and MACDUFFEE [2].

To define these formally, we proceed as follows. Let A be an n by n matrix, and let r be an integer between 1 and n. Denote by S_r the ensemble of sets of r distinct integers chosen from the integers $1, 2, \ldots, n$. Let s and t be two elements of S_r and let A_{st} denote the matrix formed from A by deleting all rows having indices that do not belong to s and all columns having indices that do not belong to t.

Enumerating the elements of S_r in some fixed order, $s_1, s_2, \ldots, s_m$, where $m = n!/r!\,(n-r)!$, construct the m by m matrix

$$C_r(A) = (|A_{s_i s_j}|) \,. \tag{1}$$

This is called the r-th compound or adjugate of A.

Its importance stems from the fact that the characteristic roots of $C_r(A)$ are the m expressions of the type $\lambda_1 \lambda_2 \ldots \lambda_r$ constituting the r-th elementary symmetric function of the λ_i. For an application of these matrices to the derivation of various inequalities, see RYSER [3]. Further ramifications can be found in BELLMAN [4], [5], KERNER [6], MONTROLL and WARD [7], FEYNMAN [8], and KARLIN and MCGREGOR [9].

§ 38. Positive Matrices

We turn now to the study of entirely different classes of matrices, which also enjoy an ordering relation.

A square matrix A with the property that all its elements are positive will be called a *positive* matrix. These matrices were introduced by PERRON [1], in connection with his thesis on the multidimensional continued fractions of JACOBI. He demonstrated the following fundamental result.

Theorem 26. *A positive matrix A possesses a unique characteristic root of largest absolute value. This root, which we shall call $p(A)$, is positive and has associated with it a positive characteristic vector, unique up to a multiplicative factor.*

FROBENIUS [2] weakened the restriction on the a_{ij} from positivity to nonnegativity and made an intensive study of this wider class of matrices. Here we shall study only positive matrices. A particularly important class of nonnegative matrices is the class of Markoff matrices or stochastic matrices of probability theory, determined by the condition that the a_{ij} represent transition probabilities,

$$\sum_{i=1}^{n} a_{ij} = 1 \,, \quad j = 1, 2, \ldots, n \,. \tag{1}$$

Theorem 26 can be demonstrated in a variety of ways ranging from methods drawn from the theory of differential equations by HARTMAN and WINTNER [3], and methods drawn purely from algebra by BRAUER [4], GANTMACHER and KREIN [5], and PERRON [1], to fixed-point techniques by ALEXANDROFF and HOPF [6], KY FAN [7], BIRKHOFF [8], and BELLMAN [9]. The proof presented below uses the fundamental ideas of the general theory of positive operators. This proof, which was communicated to us by BOHNENBLUST, is contained in BELLMAN and DANSKIN [10].

For relations of positive matrices with the branching processes of mathematical physics, see BELLMAN and HARRIS [11], and HARRIS [12]. For economic connections, see VON NEUMANN [13], WALD [14], LEONTIEFF [42.1], MORGENSTERN [42.2], ARROW and NERLOVE [42.3], and DORFMAN, SAMUELSON, and SOLOW [42.4]. Further, see KARLIN [15], DEBREU and HERSTEIN [16], SAMUELSON [17], KREIN and RUTMAN [18], and MEWBORN [19].

§ 39. Variational Characterization of $p(A)$

In this section we shall discuss a representation for $p(A)$ as the solution of a variational problem. This representation will permit us to derive some basic properties of $p(A)$ in a routine fashion.

Theorem 27. *Let A be a positive matrix, let $S(A)$ be the set of nonnegative λ for which there exist nonnegative vectors x such that $Ax \geqq \lambda x$, and let $T(A)$ be the set of positive λ for which there exist positive vectors x such that $Ax \leqq \lambda x$. Then*

$$\begin{aligned} p(A) &= \max \lambda, \quad \lambda \in S(A), \\ &= \min \lambda, \quad \lambda \in T(A). \end{aligned} \tag{1}$$

Alternatively,

$$\begin{aligned} p(A) &= \max_x \min_i \frac{\sum_{j=1}^{n} a_{ij} x_j}{x_i}, \\ &= \min_x \max_i \frac{\sum_{j=1}^{n} a_{ij} x_j}{x_i}, \end{aligned} \tag{2}$$

where, in each case, the variation is over $x_i \geqq 0$, $\sum_{i=1}^{n} x_i = 1$.

The result seems to have been discovered independently by a number of authors. The first published statement appears to be that of COLLATZ [1]; see also WIELANDT [2].

Proof. By a positive vector we mean one of which all the components are positive, with nonnegativity defined similarly. The relation $x \geqq y$ is equivalent to the statement that $x - y$ is nonnegative, or $x - y \geqq 0$, where 0 denotes the null vector. Let us normalize the vectors we consider by the condition $||x|| = \sum_{i=1}^{n} x_i = 1$, and let $||A|| = \sum_{i,j=1}^{n} a_{ij}$.

If $\lambda x \leqq A x$, we obtain

$$\lambda\, ||x|| \leqq ||A x|| \leqq ||A||\, ||x||, \tag{3}$$

where $0 \leqq \lambda \leqq ||A||$. Hence $S(A)$ is a bounded set. It is easy to see that $\lambda_0 = \sup \lambda$, $\lambda \in S(A)$, is actually a maximum. Let $x^{(0)}$ be a vector associated with λ_0; i. e., let $\lambda_0 x^{(0)} = A x^{(0)}$.

Let us now demonstrate that this relation is actually an equality. Suppose, without loss of generality, that

$$\begin{aligned} \sum_{j=1}^{n} a_{1j} x_j^{(0)} - \lambda_0 x_1^{(0)} &= d_1 > 0 , \\ \sum_{j=1}^{n} a_{kj} x_j^{(0)} - \lambda_0 x_k^{(0)} &\geqq 0 , \quad k = 2, \ldots, n . \end{aligned} \tag{4}$$

Consider the vector

$$y = x^{(0)} + \begin{pmatrix} d_1/2\lambda_0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{5}$$

It is clear that $Ay > \lambda_0 y$, since the elements of A are all positive.

This, however, contradicts the maximum property of λ_0. Thus, the assumption that d_1 satisfies the inequality $d_1 > 0$ leads to a contradiction. Similarly, we see that all the relations in (4) must be equalities.

It remains to show that $\lambda_0 = p(A)$. Let z be a characteristic vector associated with $p(A)$, so that $p(A)\, z = Az$. Let $|z|$ denote the vector having components that are the absolute values of the components of z. Then the inequality $|p(A)|\ |z| \leqq A\ |z|$ leads to the result that $|p(A)| \leqq \lambda_0$. If $|p(A)| = \lambda_0$, then the relation $|p(A)|\ |z| \leqq A\ |z|$ must be an equality. This requires that $|Az| = A\ |z|$, which means that $z = c_1 w$, where w is a nonnegative vector satisfying $Aw = p(A)$. This, however, implies that $p(A)$ is real and equal to λ_0.

To verify that the minimum definition of $p(A)$ is valid, we can either proceed along similar lines or use a more elegant argument based on the transpose matrix, or adjoint transformation, and the fact that $p(A) = p(A')$.

The proof that the associated characteristic vector is unique up to a multiplicative factor follows the preceding lines.

§ 40. A Modification due to Birkhoff and Varga

It was shown by BIRKHOFF and VARGA [1] that $p(A)$ may be written in the more elegant form

$$p(A) = \max_{\substack{x_i \geqq 0 \\ (x,x)=1}} \ \min_{\substack{y_i \geqq 0 \\ (y,y)=1}} \frac{\sum_{i,j=1}^{n} a_{ij} x_i y_j}{\sum_i x_i y_i} = \min_{\substack{y_i \geqq 0 \\ (y,y)=1}} \ \max_{\substack{x_i \geqq 0 \\ (x,x)=1}} \frac{\sum_{i,j=1}^{n} a_{ij} x_i y_j}{\sum_i x_i y_i} . \tag{1}$$

In this form, the equality of the two variational expressions is an immediate consequence of the extended min-max theorem of VON NEUMANN, which we shall discuss again in § 23 of Chapter 3.

§ 41. Some Consequences

The variational expression immediately yields the following results, which are intuitively clear from economic considerations.

Theorem 28. *If B is nonnegative and A is positive, then*

$$p(A+B) \geqq p(A). \tag{1}$$

If A_1 is a positive matrix of dimension $n-1$ formed by striking out one column and the row of like index of A, then

$$p(A_1) \leqq p(A). \tag{2}$$

Using the inequality (2), we can show that $p(A)$ is a simple root of the characteristic equation.

Another deduction is the fact that the dominant characteristic value of a matrix A is majorized by the dominant characteristic value of the matrix of the absolute values of the elements of A.

In recent years, positive matrices have assumed an important role in mathematical economics. The matrix function $p(A)$ is connected with the concept of an expanding economy (see VON NEUMANN [38.13]) and thus to the theory of games discussed in § 23 of Chapter 3.

§ 42. Input-output Matrices

Let us now consider a closely related class of square matrices defined by the conditions

$$a_{ij} > 0 \quad \text{for} \quad i \neq j\,, \quad a_{ii} \text{ real}. \tag{1}$$

A matrix satisfying these conditions we shall call an *input-output* matrix. Again, economic considerations suggest the following result. See LEONTIEFF [1], MORGENSTERN [2], ARROW and NERLOVE [3], and DORFMAN, SAMUELSON, and SOLOW [4].

Theorem 29. *If A is an input-output matrix, then A possesses a characteristic root, $r(A)$, with largest real part; this root $r(A)$ is real. The associated characteristic vector is positive, unique up to a multiplicative factor.*

Proof. We shall obtain this result as a limiting form of Theorem 27. Consider the matrix

$$e^{\delta A} = I + \delta A + \cdots, \tag{2}$$

which is positive for small positive δ. It is clear that

$$p(e^{\delta A}) = e^{\delta r(A)}. \tag{3}$$

Since this is true for a range of values of δ, $r(A)$ must be real; and $r(A) > 0$ since $p(e^{\delta A}) > 1$.

To obtain a variational characterization of $r(A)$, let us employ (39.2). We have

$$e^{\delta r(A)} = \max_x \min_i \frac{x_i + \delta \sum_{j=1}^{n} a_{ij} x_j}{x_i} + O(\delta^2) . \tag{4}$$

Hence, letting $\delta \to 0$, we obtain

$$r(A) = \max_x \min_i \frac{\sum_{j=1}^{n} a_{ij} x_j}{x_i} , \tag{5}$$

from which the conclusion of the theorem is immediate.

From this, it is easily seen that

$$r(A + B) \geqq r(A) \tag{6}$$

if A is an input-output matrix and B is a positive matrix.

§ 43. Discussion

That $p(A)$ satisfies the variational equation given in (39.2) is a consequence of the fact that the orthant $x_i \geqq 0$ is transformed into a subregion of itself by the transformation A when A is a positive matrix. Similarly, as we shall see in § 6 of Chapter 4, there is an invariant transformation associated with input-output matrices. Namely, a necessary and sufficient condition that the solution of the vector-matrix equation

$$\frac{dx}{dt} = Ax, \; x(0) = c , \tag{1}$$

be nonnegative for $t \geqq 0$, whenever $c \geqq 0$, is that A be an input-output matrix. See Bellman, Glicksberg, and Gross [1], and Arrow [2].

§ 44. Extensions

The foregoing results concerning positive matrices can be extended in a number of ways. In the first place, we have the results of Ky Fan [1] concerning the existence and characterization of values of satisfying the n equations

$$g_i(x_1, x_2, \ldots, x_n) = \lambda h_i(x_1, x_2, \ldots, x_n) , \quad i = 1, 2, \ldots, n , \tag{1}$$

where g_i and h_i are continuous functions defined on the simplex S formed by all points $(x_1, x_2, \ldots, x_n)$ with $x_i \geqq 0$ and $\Sigma x_i = 1$, the g_i are concave and > 0 on S, and the h_i convex on S, with $h_i \leqq 0$ for $x_i = 0$.

In the second place, we have the results from dynamic programming, in particular, the theory of MARKOVIAN decision processes; see BELLMAN [38.9] and HOWARD [2] concerning the solutions of the equations

$$\max_q \left[\sum_{j=1}^{n} a_{ij}(q)\, x_j\right] = x_i, \quad i = 1, 2, \ldots, n. \tag{2}$$

§ 45. Matrices and Hyperbolic Equations

In [1], P. LAX proved the following result:

Theorem 30. *Let $\{X\}$ be a linear space of n-dimensional matrices over the reals with the property that every X has only real characteristic values. Then $\lambda_{\max}(X)$, the largest characteristic value, is a convex function of X, and $\lambda_{\min}(X)$, the smallest characteristic function, is a concave function of X.*

In addition, LAX gave a representation of $\lambda_{\max}(X)$ as a supremum over linear functions. Since this representation and his proof depend heavily on the theory of linear partial differential equations of hyperbolic type, we shall omit any further details; but see WIELANDT [3] and GERSTENHABER [4].

A simple direct proof of the monotonicity of the characteristic roots of $\lambda(sX_1 + X_2)$ as s increases may be found in WEINBERGER [2]. GÅRDING pointed out that this result and the corresponding result for characteristic roots of symmetric and hermitian matrices are particular cases of his general theorem for hyperbolic polynomials; see [36.1] and [38.1] in Chapter 1.

§ 46. Nonvanishing of Determinants and the Location of Characteristic Values

Equally as important as the problem of determining upper bounds for $|A|$ is the problem of determining lower bounds, i. e., upper bounds for $|A^{-1}|$. Alternatively, we may wish merely to find conditions which ensure that $|A|$ is nonzero.

Perhaps the best-known result of this type is the following:

Theorem 31. *If*

$$|a_{ii}| > \sum_{k \neq i} |a_{ik}|, \quad i = 1, 2, \ldots, n, \tag{1}$$

then $|A| \neq 0$.

A most interesting account of this oft-discovered result, together with many other results, is contained in the expository paper by TAUSSKY [1]; see also TAUSSKY [2].

Closely related is the problem of determining regions of the λ plane that must contain characteristic values of A. A discussion of some of the extensive work on this question may be found in PARKER [3], BRAUER [4], KY FAN [5], and HAYNESWORTH [6]. Additional related results appear in OSTROWSKI [7], SLEPIAN and WEINBERG [8], [9], and KY FAN [10].

A striking consequence of the preceding theorem is the fact that the characteristic values of A are contained in the circles determined by the relations of GERSGORIN

$$|\lambda - a_{ii}| = \sum_{k \neq i} |a_{ik}| , \quad i = 1, 2, \ldots, n . \tag{2}$$

§ 47. Monotone Matrix Functions in the Sense of Loewner

Suppose we write $A \geqq B$ whenever A and B are real symmetric matrices and $A - B$ is nonnegative definite. It is easy to show by means of simple examples that $A \geqq B$ does not necessarily imply that $A^2 \geqq B^2$.

The problem of investigating the class of functions $f(z)$ with the property that $A \geqq B$ implies that $f(A) \geqq f(B)$ was first considered by LOEWNER [1], [2]. The class of functions enjoying this property for matrices of all dimensions turns out to play a prominent role in mathematical physics and in electrical network theory, where they crop up under the name of "positive real" functions.

It is easy to see that $c_1 z + c_2$, where $c_1 > 0$, is a function of the required type, but not quite as easy to see that $-z^{-1}$ is a function with an equivalent property. Namely, if $A \geqq B > 0$, then $A^{-1} \leqq B^{-1}$. We can establish this by using the quasi-linearization technique. We first demonstrate that

$$-(y, A^{-1} y) = \min_x \left[(x, A x) - 2 (x, y)\right] , \tag{1}$$

provided that $A > 0$. The minimum occurs at the value of x given by $A x = y$, whence (1) follows readily; the desired result is a consequence of (1).

It turns out that the two functions $c_1 z + c_2$ and $-z^{-1}$ are essentially the generators of the class of monotone matrix functions. See the papers by LOEWNER referred to above, and BENDAT and SHERMAN [3], where extensions to infinite-dimensional operators and other questions are discussed. Further references and discussion may be found there, and in KRAUS [4], DOBSCH [5], WIGNER [6], LOEWNER [7], DUFFIN [8], ANDERSON [9], and LANE and THOMAS [10].

§ 48. Variation-diminishing Transformations

A natural extension of a positive transformation — which is to say, a transformation that preserves the property of no variation in sign — is that of a variation-diminishing transformation. An excellent exposition of the range of problems centering about this idea is given in SCHOENBERG [1], where many other references may be found; see also his earlier paper [2], and MAIRHUBER, SCHOENBERG, and WILLIAMSON [10].

Alternatively, the study of linear vibrating circuits led GANTMACHER and KREIN to the development of "matrices complétement nonnegatives"; see [5] and [38.5]. This concept was previously developed by KELLOGG [3], [4] in connection with integral equations. The theory of convolution transforms, initiated by HIRSCHMAN and WIDDER [6], also leads naturally to the study of variation-diminishing transformations. From another direction, namely the field of Markov processes, KARLIN and McGREGOR [8] were led to further results; see also KARLIN [9]. Remarkably, there are many geometric connections; see SCHOENBERG [7]. Once again we must restrict our treatment to these sketchy comments since any adequate description of this field would require a monograph on its own.

§ 49. Domains of Positivity

Closely related to positive definite matrices is the notion of the *domain of positivity* associated with a given matrix A. We say that R is such a region if $x, y \in R$ implies that $(x, A y) \geqq 0$. Problems arising from this concept have been studied by KOECHER [1], [2].

In what has preceded, we have observed that we can generalize the concept of a positive transformation by requiring that T be a transformation that preserves an order relation, i. e.,

$$x \geqq y \rightarrow T x \geqq T y . \tag{1}$$

Another way of "inducing" positivity — which is to say, carrying over the property from one space to another — is the following. Let x and y be elements of two spaces S_1 and S_2, respectively, with the property that a scalar inner product, (x, y), is defined whenever $x \in S_1$, $y \in S_2$.

Then, provided we have a notion of positivity for all $x \in S_1$, we can define a "positive" element y of S_2 by the requirement that

$$(x, y) \geqq 0 \tag{2}$$

for *all* positive $x \in S_1$. This idea is a particular case of the main theme of the following Chapter 3.

As a particular instance, it is well known that a necessary and sufficient condition that a symmetric matrix B be positive definite is that the inequality

$$\operatorname{tr}(A B) \geqq 0 \tag{3}$$

hold for *all* positive definite matrices A. In this case, A and B are elements of the same space and no new positivity concept is generated.

Our pursuit of extensions of the notion of positivity to the vast domain of matrices has thus led us far, yet much has been left unsaid. The interested reader might consult BELLMAN [2.1] regarding stability matrices, PHILLIPS [3] regarding dissipative operators, and WIENER and MASANI [4], HELSON and LOWDENSLAGER [5], and MARSHALL and OLKIN [6], [7] regarding statistical and probabilistic considerations.

Bibliographical Notes

§ 2. An introduction to the theory of symmetric matrices, containing the results of the text and many additional discussions, may be found in the first nine chapters of

1. BELLMAN, R.: Introduction to matrix analysis. New York: McGraw-Hill Book Co., Inc. 1960.

§ 5.

1. OSTROWSKI, A., and O. TAUSSKY: On the variation of the determinant of a positive definite matrix. Nederl. Akad. Wet. Proc. (A) **54**, 383—385 (1951).

§ 6. The original results is due of ANDREIEF; see

1. ANDREIEF, C.: Note sur une relation entre les intégrales définies des products des fonctions. Mem. soc. sci. Bordeaux (3) **2**, 1—14 (1883).

This is the continuous analogue of the Lagrange formula. See also

2. DE BRUIJN, N. G.: On some multiple integrals involving determinants. J. Ind. Math. Soc. **19**, 133—151 (1955).

The question of representation as a sum of squares is intimately connected with the algebra of quadratic and higher-order forms permitting composition. For a recent exposition, see

3. JACOBSON, N.: Composition algebras and their automorphisms. Rend. Circ. Math. Palermo (2) **7**, 1—25 (1958),

where reference to earlier work by HURWITZ and others may be found. For an earlier exposition along different lines, see

4. MACDUFFEE, C. C.: On the composition of algebraic forms of higher degree. Bull. Am. Math. Soc. **51**, 198—211 (1945).

More general inequalities can be obtained from symmetrizable matrices, a generalization of symmetric matrices introduced in

5. KOLMOGOROV, A.: Zur Theorie der Markoffschen Ketten. Math. Ann. **112**, 155—160 (1936).
6. EVERITT, W. N.: Inequalities for Gram determinants. Quart. J. Math. (2) **8**, 191—196 (1957).
7. MOFFERT, C. F.: On the Gram determinant. Quart. J. Math. **10**, 161—164 (1959).
8. DE BRUIJN, N. G.: Inequalities concerning minors and eigenvalues. Nieuw. Arch. Wiskunde (3) **4**, 18—35 (1956).

This last reference [8] is an excellent expository paper in which determinantal development techniques are used to provide a unified method for dealing with a number of the inequalities of this chapter. This paper also contains some new results and a discussion of the theorems of KARAMATA and OSTROWSKI, presented in §§ 28—31 of Chapter 1.

§ 8.

1. TAUSSKY, O.: Bibliography on bounds for characteristic roots of finite matrices. National Bureau of Standards, Washington, D. C. 1951.
2. BELLMAN, R., and A. HOFFMAN: A note on an inequality of OSTROWSKI and TAUSSKY. Arkiv. Math. **5**, 123—127 (1954).

§ 9.

1. FAN, K.: On a theorem of Weyl concerning eigenvalues of linear transformations — I. Proc. Nat. Acad. Sci. USA **35**, 652—655 (1949); — II. Proc. Nat Acad. Sci. USA **36**, 31—35 (1950).

§ 10.

1. BECKENBACH, E. F.: An inequality for definite hermitian determinants. Bull. Am. Math. Soc. **35**, 325—329 (1929).

§ 11.

1. HADAMARD, J.: Resolution d'une question relative aux determinants. Bull. Sci. Math. **2**, 240—248 (1893).

This result of HADAMARD has attracted, and continues to attract, considerable attention. There are perhaps a hundred proofs available in published and unpublished form. Hadamard reproached himself for not developing the Fredholm theory of integral equations, once having obtained this result; see

2. HADAMARD, J.: The psychology of invention in the mathematical field. Princeton, N. J.: Princeton University Press, 1949.
3. BELLMAN, R.: Notes on matrix theory — II. Am. Math. Monthly **60**, 174—175 (1953).

A particularly interesting way of establishing the Hadamard inequality is given in

4. MARCUS, M.: Some properties and applications of doubly stochastic matrices. Am. Math. Monthly **67**, 215—221 (1960).

The author shows that the same general techniques suffice to establish the Ky Fan inequality of § 32 and the Minkowski inequality of § 22 in this chapter.

5. OSTROWSKI, A. M.: On some metrical properties of operator matrices and matrices partitioned into blocks. J. Math. Anal. Appl. **2** (1961), to appear.

§ 12. For various types of extensions, see

1. FISCHER, E.: Über den Hadamardsche Determinantensatz. Arch. Math. Phys. (3) **13**, 32—40 (1908).
2. WILLIAMSON, J.: Note on Hadamard's determinant theorem. Bull. Am. Math. Soc. **53**, 608—613 (1947).
3. SCHUR, I.: Über endliche Gruppen und Hermitesche Formen. Math. Z. **1**, 184—207 (1918).
4. BUSH, K. A., and I. OLKIN: Extrema of quadratic forms with applications to statistics. Biometrika **46**, 483—486 (1959).
5. MIRSKY, L.: On a generalization of Hadamard's determinantal inequality due to SZÁSZ. Arch. Math. **8**, 274—275 (1957).

§ 13.

1. BELLMAN, R.: Hermitian matrices and representation theorems. Duke Math. J. **26**, 485—490 (1959).

This paper partially extends results of HUA,

2. HUA, L. K.: Inequalities involving determinants. Acta Math. Sinica **5**, 463—470 (1955),

where an entirely different type of representation of a determinant, based on group representation theory, is used.

§ 15. The generalized Euler integral was discovered independently by INGHAM and SIEGEL. See

1. INGHAM, A. E.: An integral which occurs in statistics. Proc. Cambridge Phil. Soc. **29**, 271—276 (1933),
2. SIEGEL, C. L.: Über die analytische Theorie der quadratischen Formen. Ann. of Math. **36**, 527—606 (1935).

Generalizations were given by

3. OLKIN, I.: A class of integral identities with matrix argument. Duke Math. J. **26**, 207—213 (1959).
4. GÅRDING, L.: The solution of CAUCHY's problem for two totally hyperbolic linear differential equations by means of Riesz integrals. Ann. of Math. **48**, 785—826 (1947).
5. AITKEN, A. C.: On the Wishart distribution in statistics. Biometrika **36**, 59—62 (1949).
6. BELLMAN, R.: A generalization of some integral identities due to INGHAM and SIEGEL. Duke Math. J. **24**, 571—578 (1956).

The integral discussed by SIEGEL plays an important role in his theory of matrix modular functions. Combined with the Poisson summation formula, it yields an elegant matrix version of an identity due to LIPSCHITZ. Similarly, the generalized integral in BELLMAN [6] can be used to obtain a further generalization of the Siegel result.

§ 17.

1. BERGSTROM, H.: A triangle inequality for matrices. Den Elfte Skandinaviski Matematiker-kongress, Trondheim, 1949. Oslo: Johan Grundt Tanums Forlag 1952.
2. BELLMAN, R.: Notes on matrix theory — IV: an inequality due to Bergstrom. Am. Math. Monthly **62**, 172—173 (1955).

§ 18. This generalization is an extension of a partial result based on the integrals of § 15, given in

1. BELLMAN, R.: Notes on matrix theory — IX. Am. Math. Monthly **64**, 189—191 (1957).

§ 20. This result is given in [18.1]. See also § 44 of Chapter 1 for equivalent and related results.

§ 21. A large number of similar, but more complicated, results can be obtained for the other basic functions of A (the coefficients in the expansion of the characteristic polynomial $|\lambda I + A|$), by means of the results of GÅRDING on hyperbolic polynomials cited in §§ 36—38 of Chapter 1. From these representations, or otherwise, a number of further inequalities can be obtained.

See, in Chapter 1, the papers [36.1] and [38.1] by GÅRDING, [38.2] by CHERN, and also the paper [8.2] by LORCH, where the connection between these results and the "mixed volume" of MINKOWSKI is given. Despite the many points of overlap, we have avoided any discussion of such matters here, reserving them for our second volume on inequalities.

An interesting extension of Theorem 14 is due to O. TAUSSKY, namely,

$$|A^{(k)}|^{1/m_k} = \min_{|B|=1} \frac{\operatorname{tr}(A^{(k)} B^{(k)})}{\binom{n}{k}} .$$

§ 22.

1. HAYNESWORTH, E. V.: Note on bounds for certain determinants. Duke Math. J. **24**, 313—320 (1957).
2. HAYNESWORTH, E. V.: Bounds for determinants with positive diagonals. Trans. Am. Math. Soc. **96**, 395—413 (1960).

These papers contain results of OSTROWSKI, PRICE, SCHNEIDER, and BRENNER, and many additional references.

§ 23.

1. FAN, K.: Some inequalities concerning positive-definite hermitian matrices. Proc. Cambridge Phil. Soc. **51**, 414—421 (1955).
2. FAN, K.: Problem 4786. Am. Math. Monthly **65**, 289 (1958).

3. MIRSKY, L.: Maximum principles in matrix theory. Proc. Glasgow Math. Assoc. **4**, 34—37 (1958).

§ 24.

1. OPPENHEIM, A.: Inequalities connected with definite hermitian forms. J. London Math. Soc. **5**, 114—119 (1930).
2. OPPENHEIM, A.: Inequalities connected with definite hermitian forms, II. Am. Math. Monthly **61**, 463—466 (1954).

§ 25. See R. BELLMAN [2.1] for further discussion of the contents of §§ 25, 26.

§ 26. The results of Courant for continuous operators may be found in

1. COURANT, R., and D. HILBERT: Methods of mathematical physics. New York: Interscience Publishers, Inc. 1953.

The original result of FISCHER is given in

2. FISCHER, E.: Über quadratische Formen mit reellen Koeffizienten. Monatsh. Math. Physik **16**, 234—249 (1905).

§ 27.

1. BELLMAN, R., I. GLICKSBERG and O. GROSS: Notes on matrix theory — VI. Am. Math. Monthly **62**, 571—572 (1955).

§ 28.

1. FAN, K.: Problem 4430. Am. Math. Monthly **58**, 194—195 (1951).

§ 29. See Part II of [9.1].

§ 30.

1. WEYL, H.: Inequalities between the two kinds of eigenvalues of a linear transformation. Proc. Nat. Acad. Sci. USA **35**, 408—411 (1949).
2. FAN, K.: Maximum properties and inequalities for the eigenvalues of completely continuous operators. Proc. Nat. Acad. Sci. USA **37**, 760—766 (1951).
3. LIDSKII, V. B.: The proper values of the sum and product of symmetric matrices. Doklady Akad. Nauk. USSR **75**, 769—772 (1950).
4. HORN, A.: On the singular values of a product of completely continuous operators. Proc. Nat. Acad. Sci. USA **36**, 374—375 (1950).
5. WIELANDT, H.: An extremum property of sums of eigenvalues. Proc. Am. Math. Soc. **6**, 106—110 (1955).
6. AMIR-MOEZ, ALI R.: Extreme properties of eigenvalues of a hermitian transformation and singular values of the sum and product of a linear transformation. Duke Math. J. **23**, 463—477 (1956).
7. MIRSKY, L.: On a convex set of matrices. Arch. Math. **10**, 88—92 (1959).

§ 31.

1. CAUCHY, A.: Sur l'équation à l'aide de laquelle in détermine les inegalités seculaires des mouvements des planetes. Oeuvres Complètes II^{e} serie, IX, 174—195. Paris: Debure frères 1821.
2. POINCARÉ, H.: Sur les équations aux derivées partielles de la physique mathématiques. Am. J. Math. **12**, 211—294 (1890).

See also

3. PÓLYA, G.: Estimates for eigenvalues. Studies in Mathematics and Mechanics. New York: Academic Press, Inc. 1954.
4. HAMBURGER, H. L., and M. E. GRIMSHAW: Linear transformations in n-dimensional vector space. Cambridge: Cambridge University Press 1951.

A simpler formulation of the separation theorem is: "If A is a real symmetric matrix of order n and B a matrix obtained as a principal submatrix of A of order k, then $\lambda_i(B) \leqq \lambda_i(A)$, $i = 1, 2, \ldots, k$, $\lambda_{n-j}(B) \geqq \lambda_{n-j}(A)$, $j = 0, 1, \ldots, k-1$." (We owe this formulation to KY FAN.)

Reference [3] contains some interesting remarks concerning the equivalence of the separation and the min-max theorem of FISCHER, as well as some applications.

§ 32.

1. FAN, K.: Problem 4429. Am. Math. Monthly **58**, 194 (1951).

§ 33. We owe this discussion to KY FAN.

§ 34. This result is contained in Part I of [9.1].

1. MIRSKY, L.: An inequality for positive definite matrices. Am. Math. Monthly **62**, 428–430 (1955).

§ 35.

1. BELLMAN, R.: Note on matrix theory — multiplicative properties from additive properties. Am. Math. Monthly **65**, 693–694 (1958).

§ 36.

1. MARCUS, M., and J. L. MCGREGOR: Extremal properties of hermitian matrices. Canad. J. Math. **8**, 524–531 (1956).
2. MARCUS, M., and B. N. MOYLS: Extreme value properties of hermitian matrices. Dept. of Math., Univ. of British Columbia, Vancouver, Canada 1956.
3. MARCUS, M., B. N. MOYLS and R. WESTWICK: Some extreme value results for indefinite hermitian matrices, II. Illinois J. Math. **2**, 408–414 (1958).
4. MARCUS, M., B. N. MOYLS and R. WESTWICK: Extremal properties of hermitian matrices. Canad. J. Math. (to appear).

§ 37. The elegant idea of studying various classes of associated matrices was introduced by SCHUR in his thesis,

1. SCHUR, I.: Über eine Klasse von Matrizen die sich einer gegebenen Matrizc zuordnen lassen. Dissertation, Berlin 1901.

For a detailed discussion and many further references, see

2. MACDUFFEE, C. C.: The theory of matrices. New York: Chelsea Publishing Co. 1946.
3. RYSER, H. J.: Inequalities of compound and induced matrices with applications to combinatorial analysis. Illinois J. Math. **2**, 240–253 (1958).

It is quite remarkable that the Kronecker products of matrices and the compound matrices of Schur arise in a natural fashion not only in group theory and geometry (see, for example, LORCH [8.2] in Chapter 1) as one would expect, but also in various parts of analysis. Thus, the Kronecker product arises in a discussion of stability theory, for which see

4. BELLMAN, R.: Kronecker products and the second method of LYAPUNOV. Math. Nachr. **20**, 17–19 (1959),

and also in the study of stochastic processes, for which see

5. BELLMAN, R.: Limit theorems for non-commutative operations — I. Duke Math. J. **21**, 491–500 (1954).
6. KERNER, E. H.: The band structure of mixed linear lattices. Proc. Phys. Soc. **69**, 234–244 (1956).

The SCHUR compound matrices, the elements of which are generalized Plücker coordinates, arise naturally in the study of Markoff processes involving k particles, no two of which are allowed to occupy the same state at the same time; see

7. MONTROLL, E. W., and J. C. WARD: Quantum statistics of an electron gas. Bell Telephone Monograph 3036. New York: Bell Telephone Co. 1959.
8. FEYNMAN, R. P.: The theory of positrons. Space-time approach to electrodynamics. Phys. Rev. **76**, 749—759, 769—789 (1949).
9. KARLIN, S., and J. MCGREGOR: Coincidence probabilities. Department of Statistics, Tech. Report No. 8. Stanford Calif.: Stanford University Press 1959.

One theme that we have not exploited is the natural positivity, or nonnegativity, associated with probabilities. It turns out that this is intimately connected with the variation-diminishing transformations of SCHOENBERG to which we shall briefly refer in § 48.

This technique could also be employed in Chapter 4 in connection with our discussion of the nonnegativity of GREEN's functions. We decided to omit any treatment of this very powerful technique in the present volume on the grounds that the required preliminary discussions would take us too far astray.

§ 38. It is only in recent years that proper attribution of the fundamental result on positive matrices has been made to PERRON:

1. PERRON, O.: Zur Theorie der Matrizen. Math. Ann. **64**, 248–263 (1907).

A number of texts and papers still credit the basic result to FROBENIUS, who extended and developed the original theorem of PERRON; see

2. FROBENIUS, G.: Über Matrizen aus nicht-negativen Elementen. Sitzber. preuß. Akad. Wiss. **1912**, 456–477.

Various proofs of Theorem 26 have been given; see

3. HARTMAN, P., and A. WINTNER: Linear differential equations and difference equations with monotone solutions. Am. J. Math. **75**, 731–743 (1953).
4. BRAUER, A.: A new proof of theorems of PERRON and FROBENIUS on non-negative matrices — I: positive matrices. Duke Math. J. **24**, 367–378 (1957).
5. GANTMACHER, V., and M. KREIN: Sur les matrices complétement non-negatives and oscillatoires. Comp. Math. **4**, 445–476 (1937).
6. ALEXANDROFF, P., and H. HOPF: Topologie, I. Berlin: J. Springer Verlag 1935.
7. FAN KY: Topological proofs for certain theorems on matrices with nonnegative elements. Monatsh. Math. **62**, 219–237 (1958).
8. BIRKHOFF, G.: Extensions of JENTZSCH's theorem. Trans. Am. Math. Soc. **85**, 219–227 (1957).
9. BELLMAN, R.: On a quasi-linear equation. Canad. J. Math. **8**, 198–202 (1956).
10. BELLMAN, R., and J. DANSKIN: A survey of the mathematical theory of time lag, retarded control, and hereditary processes. Santa Monica, Calif.: The RAND Corporation, Report R-256, 1954.

The theory of positive operators has come into prominence as a result of at least three different influences, stemming from internal mathematical considerations, from mathematical physics, and from mathematical economics. Apart from the work of PERRON and FROBENIUS, one is led to the study of nonnegative matrices by means of invariance considerations. As the orthogonal matrices preserve the orthogonal form (x, x), the unitary matrices preserve the form $(x, \bar{x})$, and the Markoff matrices preserve the probability region $x_i \geqq 0$, $\sum_{i=1}^{n} x_i = 1$, so the nonnegative matrices preserve the region defined by $x_i \geqq 0$, the positive orthant. Examining the situation from this point of view, we see that many interesting classes of matrices remain to be studied, since there are many other important types of "positivity."

From the standpoint of mathematical physics, the theory of branching processes — covering such diverse phenomena as neutron fission, cosmic-ray cascades, and biological growth and mutation — presents a number of problems requiring a general theory of positive operators. For a discussion of such matters, see

11. BELLMAN, R., and T. E. HARRIS: On the theory of age-dependent stochastic branching processes. Proc. Nat. Acad. Sci. USA **34**, 601–604 (1948).
12. HARRIS, T. E.: Branching processes. Ergebnisse der Mathematik. Berlin-Göttingen-Heidelberg: J. Springer Verlag 1961.

Finally, the study of the expanding economy initiated by VON NEUMANN,

13. VON NEUMANN, J.: Über ein ökonomisches Gleichungssystem und die Verallgemeinerung des Brouwerschen Fixpunktsatzes. Vienna: Ergebnisse eines Mathematischen Kolloquiums, vol. 8, 73–83 (1937),

and contributed to by WALD, MORGENSTERN, KEMENY, THOMPSON, and others, for example, in

14. WALD, A.: Über die Productionsgleichungen der ökonomischen Wertlehre. II. Vienna: Ergebnisse eines Mathematischen Kolloquiums, vol. **7**, 1–6 (1936),

and also the input-output models of LEONTIEFF, to which we shall refer in § 42, have shown the great importance of these matrices in the theory of mathematical economics. The "input-output" matrices were first studied by MINKOWSKI in a different connection.

A detailed account of work in this field may be found in

15. KARLIN, S.: Mathematical methods and theory in games, programming and economics, 1, 2. Reading, Mass.: Addison-Wesley Publishing Company, Inc. 1959.
16. DEBREU, G., and I. N. HERSTEIN: Non-negative square matrices. Econometrica **21**, 597–607 (1953).
17. SAMUELSON, H.: On the Perron-Frobenius theorem. Michigan Math. J. **4**, 57–59 (1957).
18. KREIN, M. G., and M. A. RUTMAN: Linear operators leaving invariant a cone in a Banach space. Am. Math. Soc. Translation, No. 26. New York: American Mathematical Society 1950.
19. MEWBORN, A. C.: Generalizations of some theorems on positive matrices to completely continuous linear transformations in a normed linear space. Duke Math. J. **27**, 273–277 (1960).

§ 39. As indicated in the text, the variational characterization of $p(A)$ has been rediscovered many times. F. J. MURRAY, in an unpublished paper on the Ising problem, uses this result and attributes it to VON NEUMANN. The result is implied in

1. COLLATZ, L.: Einschließungssatz für die charakteristischen Zahlen von Matrizen. Math. Z. **48**, 221–226 (1946),

and first stated precisely in

2. WIELANDT, H.: Unzerlegbare, nicht negative Matrizen. Math. Z. **52**, 642–648 (1950).

The proof given here was furnished by BOHNENBLUST in connection with a problem arising in multidimensional branching processes considered by BELLMAN and HARRIS; see [38.11].

§ 40.

1. BIRKHOFF, G., and R. S. VARGA: Reactor criticality and nonnegative matrices. J. Ind. Appl. Math. **6**, 354–377 (1958).

The min-max theorem of VON NEUMANN, the keystone of the BOREL and VON NEUMANN theory of games, will be discussed in § 23 of Chapter 3.

§ 41. See the references [38.13] and [38.14] to VON NEUMANN and WALD.

§ 42. For a detailed discussion of input-output matrices, see KARLIN [38.15], where many additional references can be found. Particularly, we cite the following works:

1. LEONTIEFF, W. W.: The structure of American economy. Cambridge, Mass.: Harvard University Press, 1941.
2. MORGENSTERN, O. (editor): Economic activity analysis. New York: John Wiley and Sons 1954.
3. ARROW, K. J., and M. NERLOVE: A note on expectation and stability. Econometrica **26**, 297–305 (1958).

In each of these, many additional results and references will be found. For a general discussion of the closely related topic of linear programming (mentioned again in Chapter 3, below), and its place in mathematical economics, see

4. DORFMAN, R., H. SAMUELSON and R. SOLOW: Linear programming and economic analysis. New York: McGraw-Hill Book Co., Inc. 1958.

§ 43. This result was given in

1. BELLMAN, R., I. GLICKSBERG and O. GROSS: On some variational problems occurring in the theory of dynamic programming. Rend. Circ. Mat. Palermo (2) **3**, 1—35 (1954).

For an ingenious proof of the negativity of $-A^{-1}\mathrm{b}$ based on the nonnegativity of the solutions of $du/dt = Au + \mathrm{b}$, see

2. ARROW, K. J.: Price-quantity adjustments in multiple markets with rising demands. Mathematical models in the social sciences. Stanford, Calif.: Stanford University Press 1959.

§ 44.

1. FAN, K.: On the equilibrium value of a system of convex and concave functions. Math. Z. **70**, 271—280 (1958).
2. HOWARD, R.: Discrete dynamic programming. New York: John Wiley and Sons 1960.

§ 45.

1. LAX, P. D.: Differential equations, difference equations, and matrix theory. Comm. Pure Appl. Math. **10**, 175—194 (1958).
2. WEINBERGER, H. F.: Remarks on the preceding paper of LAX. Comm. Pure Appl. Math. **10**, 196—197 (1958).

See the papers by GÅRDING and CHERN referred to in §§ 36—38 of Chapter 1.

For connections between the theory of symmetric matrices and the results of LAX, see

3. WIELANDT, H.: Lineare Scharen von Matrizen mit reellen Eigenwerten. Math. Z. **53**, 219—225 (1950),

and the extension in

4. GERSTENHABER, M.: Note on a theorem of WIELANDT. Math. Z. **71**, 141—142 (1959).

§ 46.

1. TAUSSKY, O.: A recurring theorem on determinants. Am. Math. Monthly **56**, 672—676 (1949).
2. TAUSSKY, O.: Bibliography of bounds for characteristic roots of finite matrices, Nat. Bur. Standards Report. Washington, D. C.: National Bureau of Standards 1951.
3. PARKER, W. W.: Characteristic roots and fields of value of a matrix. Bull. Am. Math. Soc. **57**, 103—108 (1951).
4. BRAUER, A.: Limits for the characteristic roots of a matrix. I, II, III. Duke Math. J. **13**, **14**, **15**, 387—395, 21—26, 871—877 (1946, 1947, 1948).
5. FAN, K.: Note on circular disks containing the eigenvalues of a matrix. Duke Math. J. **25**, 441—445 (1958).
6. HAYNSWORTH, E. V.: Bounds for determinants with dominant main diagonal. Duke Math. J. **20**, 199—209 (1953).

A real matrix A for which $|A_{ii}| > \sum_{j \neq i} |a_{ij}|$ is sometimes called a *Hadamard matrix*. These matrices are closely related to matrices studied by MINKOWSKI, to the M-matrices of OSTROWSKI,

7. OSTROWSKI, A.: Note on bounds on determinants. Duke Math. J. **22**, 95—102 (1955),

where references to earlier work may be found, and to the dominant matrices of SLEPIAN and WEINBERG.

8. SLEPIAN, P., and L. WEINBERG: Synthesis of paramount and dominant matrices. Los Angeles, Calif.: Hughes Research Laboratories 1958.
9. SLEPIAN, P., and L. WEINBERG: Positive real matrices. J. Math. and Mech. **9**, 71—84 (1960).

See also

10. FAN, K.: Note on M-matrices. Quart. J. Math. **11**, 43—49 (1960).

Matrices of these types play an important part in the computational solution of partial differential equations via difference schemes.

Finally, let us point out that the theorem that $|A| > 0$ if $a_{ii} > \sum_{i \neq j} |a_{ij}|$ can be proved by means of the same continuity argument used in § 3. The general result is the following: If A belongs to a convex set of matrices containing the identity matrix, and if $|A| \neq 0$, then $|A|$ is actually positive.

§ 47.

1. LOEWNER, C.: Über monotone Matrix funktionen. Math. Z. **38**, 177—216 (1934).
2. — Some classes of functions defined by difference or differential inequalities. Bull. Am. Math. Soc. **56**, 308—319 (1950).
3. BENDAT, J., and S. SHERMAN: Monotone and convex operator functions. Trans. Am. Math. Soc. **79**, 58—71 (1955).
4. KRAUS, F.: Über konvexe Matrixfunktionen. Math. Z. **41**, 18—42 (1936).
5. DOBSCH, R.: Matrixfunktionen beschränkter Schwankung. Math. Z. **43**, 353—388 (1937).
6. WIGNER, E.: On a class of analytic functions from the quantum theory of collisions. Ann. of Math. **53**, 36—67 (1951).
7. LOEWNER, C.: On totally positive matrices. Math. Z. **63**, 338—340 (1955).
8. DUFFIN, R. J.: Elementary operations which generate network matrices. Proc. Am. Math. Soc. **6**, 335—339 (1955).
9. ANDERSON, T. W.: The integral of a symmetric unimodal function. Proc. Am. Math. Soc. **6**, 170—176 (1955).
10. LANE, A. M., and R. G. THOMAS: R-matrix theory of nuclear reactions. Rev. Mod. Phys. **30**, 257—352 (1958).

§ 48.

1. SCHOENBERG, I. J.: On smoothing operations and their generating functions. Bull. Am. Math. Soc. **59**, 199—230 (1953).
2. — Über variationsvermindernde lineare Transformationen. Math. Z. **32**, 321—328 (1930).

The theory of completely positive matrices of GANTMACHER and KREIN is an algebraic analogue of the Kellogg kernels in the theory of integral equations. See

3. KELLOGG, O. D.: The oscillation of functions of an orthogonal set. Am. J. Math. **38**, 1—5 (1916),
4. KELLOGG, O. D.: Orthogonal function sets arising from integral equations. Am. J. Math. **40**, 145—154 (1918),

and a series of papers on these kernels by GANTMACHER and KREIN, references to which may be found in [38.5] and in the book

5. GANTMACHER, F. R., and M. G. KREIN: Oszillationsmatrizen, Oszillationskerne und kleine Schwingungen mechanischer Systeme. Berlin: Akademie-Verlag 1960.

Further references, together with some of the principal results and applications, are given in

6. HIRSCHMAN, I. I., and D. V. WIDDER: The convolution transform. Princeton, N. J.: Princeton University Press 1955.

7. SCHOENBERG, I. J.: An isoperimetric inequality for closed curves convex in even-dimensional Euclidean spaces. Acta Math. **91**, 143—164 (1954).

For some applications of these ideas to probability and statistics, see

8. KARLIN, S., and J. L. MCGREGOR: The differential equations of birth-and-death processes and the Stieltjes moment problem. Trans. Am. Math. Soc. **85**, 489—546 (1957).

9. KARLIN, S.: Pólya-type distributions — IV. Ann. Math. Stat. **29**, 1—21 (1958).

10. MAIRHUBER, J. C., I. J. SCHOENBERG and R. E. WILLIAMSON: On variation-diminishing transformations of the circle. Rend. Circ. Mat. Palermo (2) **8**, 241—270 (1959).

§ 49.

1. KOECHER, M.: Die Geodatischen von Positivitätsbereichen. Math. Ann. **135**, 192—202 (1958).

2. KOECHER, M.: Positivitätsbereiche im R^n. Am. J. Math. **79**, 575—596 (1957).

Very little is known about "stability matrices," i. e., real matrices all the characteristic roots of which have negative real parts. Some results may be found in [2.1], where further references are given. There is, however, an extensive theory of "dissipative operators"; see

3. PHILLIPS, R. S.: On the integration of the diffusion equation with boundary conditions. Trans. Am. Math. Soc. **98**, 62—84 (1961).

A number of interesting inequalities pertaining to matrices are given in

4. WIENER, N., and P. MASANI: The prediction theory of multivariate stochastic processes — I, II. Acta Math. **58**, 111—148 (1957); **99**, 93—130 (1958).

See also

5. HELSON, H., and D. LOWDENSLAGER: Prediction theory and Fourier series in several variables. Acta Math. **99**, 165—202 (1958).

For a number of matrix inequalities derived from statistical and probabilistic considerations, see

6. MARSHALL, A., and I. OLKIN: Technical Report, No. 36, Applied Math. and Statistics. Stanford, Calif.: Stanford University Press 1960.

7. MARSHALL, A., and I. OLKIN: Multivariate Chebyshev Type Inequalities, East Lansing, Mich.: Michigan State University 1959.

Chapter 3

Moment Spaces and Resonance Theorems

§ 1. Introduction

A central idea of analysis, which can be used to connect vast fields of study that at first glance may seem quite unrelated, can be expressed in the following simple form:

"An element of a linear space S can often be characterized most readily and revealingly in terms of its interaction with a suitably chosen set of elements in a dual space S'."

This principle rather naturally finds its source in geometry, in the concept of poles and polars, the most important case of which is the point-tangent relationship. We have already used this idea in Chapter 1 in our discussion of the Minkowski-Mahler inequality, and in our repeated appeal to quasilinear representations in Chapters 1 and 2. The duality between point and line characterizations permeates algebra, geometry, and analysis. In all likelihood, it is the single most ramified concept in mathematics.

Many instances of this "projection" technique may be found in the theory of Fourier series, and more generally in the theory of orthogonal expansions. The concept reaches its full flowering, however, only in the theory of Banach spaces, and it is in the theory of Hilbert spaces that it is most powerful. Elegant applications have been made in the theory of partial differential equations, where the techniques of orthogonal projection, weak solutions, and generalized solutions play dominant roles. See BANACH [1] and ZYGMUND [2].

In this chapter we shall discuss a number of "interaction" theorems, tracing some of their common origins and their many interconnections, and exhibiting their relations to the theory of inequalities. These relations arise from the fact that in many important cases the interaction between $x \in S$ and $y \in S'$ is expressed by an inequality of the form

$$(x, y) \geqq 0\,, \tag{1}$$

with a suitable definition of the inner product.

Occasionally, as in various quasilinear representations, a more complicated expression is required.

When problems are phrased in the foregoing fashion, it is reasonable to expect that arguments borrowed on the one hand from the theory of convex sets and on the other hand from the theory of linear spaces will have a unifying and simplifying role. This is certainly the case. The method of orthogonal projection, the Banach-Steinhaus theorem and the Hahn-Banach extension theorem, and the notion of the "separating plane" all play fundamental and interchangeable parts.

It is important to note that in studying these basic questions of mathematics, the artificial boundaries of algebra, geometry, and analysis melt away. There are certain major problems and methods of solution that cover and unite all these subdomains. For an excellent exposition of the solution of a number of significant problems in analysis from this unified point of view, see ROSENBLOOM [3].

In view of the many different regions of analysis that are penetrated and illuminated by these algebraic-geometric ideas, it is equally logical and equally profitable to begin our discussion in any of a number of ways.

We shall motivate our discussion initially as a continuation of the Cauchy-Schwarz-Buniakowsky and Hölder inequalities. In place of the particular results we have previously obtained for the power sums

$$s_k = x_1^k + x_2^k + \cdots + x_k^n\,, \tag{2}$$

$k = 1, 2, \ldots,$ where the x_i are nonnegative quantities, we wish to determine a complete set of inequalities. By a complete set of inequalities, we mean a set of relationships of the form

$$R_i(s_1, s_2, \ldots, s_k) \geqq 0\,, \quad i = 1, 2, \ldots, \tag{3}$$

that are satisfied by all sequences of the form appearing in (2), and that, conversely, ensure that any sequence $\{s_k\}$ satisfying these relationships can be represented in the form given in (2); see, for example, URSELL [2.1].

This problem is, of course, fundamental in the theory of moment spaces. Our aim is not to enter this rather vast field in any depth or in any force, but rather to indicate how these questions are related to the study of inequalities, what general methods can be used to investigate them, and what types of results can be obtained. For detailed investigations of the variety of results that exist, we shall refer the reader to a number of excellent treatises and books.

It is soon seen that we may just as well consider the more general problem of determining complete sets of inequalities for general "moments",

$$m_k = (x, y_k)\,, \quad k = 1, 2, \ldots \tag{4}$$

As above, $x \in S$, $y_k \in S'$, and S and S' are two spaces for which an inner product of respective elements can be defined.

The problem formulated in this fashion furnishes an excellent setting for the introduction of simple and intuitive geometric ideas centering about the concept of convexity. After introducing these ideas, we derive, as an immediate consequence, a well-known theorem of F. RIESZ; this is a completion and extension of the inequality of BESSEL in the theory of orthogonal series, and also of the HÖLDER inequality.

Following this, we turn to the problem of determining complete sets of inequalities for the ordinary moments

$$m_k = \int_0^1 x^k \, dG(x)\,, \tag{5}$$

and for the trigonometric moments

$$m_k = \int_0^1 e^{2\pi i k x} \, dG(x)\,. \tag{6}$$

We linger long enough in this area to exhibit the connection between these problems and the problem of the representation of positive functions, "positive" in some sense or other, as squares and sums of squares. These questions, of great fascination and importance in their own right, have been briefly touched upon in Chapter 2 in connection with positive definite quadratic forms. Consequently, representations of nonnegative functions as sums of squares link the investigations of this chapter with those of Chapter 2. A surprising result is that complete sets of inequalities can be obtained directly from the determinantal inequalities characterizing positive definite quadratic forms, established in §§ 3—5 of Chapter 2.

The concept of a positive definite function, introduced by MATHIAS [4] and refined by BOCHNER [5], plays a paramount role in many parts of modern analysis, as a consequence of a fundamental representation theorem due to BOCHNER [5]; see also COOPER [6]. For some of its varied extensions, see the papers by GODEMONT [9], and CARTAN and GODEMONT [10], where work of GELFAND and RAIKOV is covered. We shall also refer the reader to some work of BOCHNER, VON NEUMANN and SCHOENBERG on Hilbert spaces and positive definite functions [15.4], [15.5], [15.6], and to work by ARONSZAJN and SMITH on reproducing kernels [16.1], [16.2]. Additional material will be found in FAN [7] and WIDDER [8].

The theory of moments has been so thoroughly and elegantly treated, in a number of easily available sources, that we shall bypass the principal questions completely. We shall present some results for trigonometric moments showing the connection between the representation theorem for a positive trigonometric polynomial and theorems of this nature. The theory is much simpler for trigonometric polynomials than for ordinary polynomials, due to the simpler representations that exist for trigonometric polynomials. See WIDDER [8], SHOHAT and TAMARKIN [11], and KARLIN and SHAPLEY [12].

In passing, we shall indicate the rather unexpected connection between this problem area and the classical results of PICARD and LANDAU concerning entire functions that omit particular values. At the conclusion of the chapter, we shall briefly indicate relations to the theory of TOEPLITZ matrices; see the recent book by GRENANDER and SZEGÖ [13.11].

An excellent survey of the origin of extensive classes of questions in moment theory within the framework of classical probability theory is contained in the expository paper by MALLOWS [13], where many further references are given. Here one will find a discussion of the determination of moment spaces connected with unimodal distributions, and so on.

As an example of the use of convexity arguments in a setting in which at first sight they do not seem to apply, we consider the problem of determining the range of the moments determined by expressions of the form

$$m_k = \int_0^1 \phi(x)\, x^k\, dx\,, \quad k = 1, 2, \ldots, \tag{7}$$

where the function $\phi(x)$ is a characteristic function of a set $S \supset [0,1]$, i. e.,

$$\begin{aligned} \phi(x) &= 1\,, \quad x \in S\,, \\ \phi(x) &= 0\,, \quad \text{otherwise.} \end{aligned} \tag{8}$$

Turning from these questions, which as mentioned above could be dwelt upon at great length, we examine an interesting theorem due to LANDAU, which completes the Hölder inequality in an appropriate sense. The content of this result is that a vector x, with components $x_1, x_2, \ldots, x_n, \ldots$, belongs to l^p, the space of sequences $\{x_n\}$ for which $\sum_n |x_n|^p < \infty$, $p > 1$, if and only if its projections on all vectors in l^q, where $q = p/(p-1)$, are finite. We recognize this, of course, as a forerunner of the elegant theorem of BANACH and STEINHAUS [1], useful in many areas of analysis. Clearly it is closely connected to the result of F. Riesz mentioned above, and presented below in § 7.

The theme of resonance theorems takes us next into the domain of linear inequalities, where an analogue of the Landau theorem is a basic discovery of MINKOWSKI. This, in turn, is a particular result in the general theory of linear inequalities, a classical discipline that in recent years has been revitalized in connection with the theory of games of BOREL and VON NEUMANN and the theory of linear programming. Here again, since so much work has been published recently, questions will be discussed in quite cursory fashion. For basic theory and many references, the reader is referred to KARLIN ([38.15] in Chapter 2) and to DORFMAN, SAMUELSON and SOLOW ([42.4] in Chapter 2).

There are many discussions of the connection between the general theory of convex functions and linear inequalities. The recent work of KY FAN [15], [16], and KY FAN, GLICKSBERG, and HOFFMAN [14] in this field draws upon both convexity arguments and the theory of linear and Banach spaces. In particular, the Hahn-Banach theorem plays a vital role, demonstrating once again the intimate relation between this result and the theorem of the separating hyperplane.

To illustrate the difference between the techniques used and the results obtainable in the theory of linear inequalities, and those of classical theory, we shall state and prove the Neyman-Pearson lemma. This result, which has its origin in the field of mathematical statistics, has recently become of importance in connection with economic and engineering control processes. See [17], and also [19.1] in Chapter 1.

We shall close the chapter with some brief remarks concerning the technique of orthogonal projection foreshadowed by ZAREMBA and developed by WEYL [18], LAX [19], and others; the connection of this method with DU BOIS REYMOND's "fundamental lemma" of the calculus of variations, for which see GRAVES [20] and BERWALD [21]; and finally the relationship between these ideas and the generalized solutions of partial differential equations, developed by BOCHNER and FRIEDRICHS and systematized in the form of the theory of distributions by L. SCHWARTZ [22].

Throughout, our aim will be to spotlight fundamental concepts and to illustrate their interrelations and farreaching applications.

§ 2. Moments

Consider the expression

$$s_k = x_1^k + x_2^k + \cdots + x_n^k, \tag{1}$$

$k = 1, 2, \ldots$, where the x_i are nonnegative quantities. When we write

$$s_{k+l} = x_1^k\, x_1^l + x_2^k\, x_2^l + \cdots + x_n^k\, x_n^l, \tag{2}$$

the Cauchy-Schwarz inequality yields the result

$$s_{k+l}^2 \leqq s_{2k} s_{2l}. \tag{3}$$

Furthermore, we have the additional set of relationships

$$s_1 \geqq s_2^{1/2} \geqq \cdots \geqq s_n^{1/n} \geqq \cdots, \tag{4}$$

and it is clear that many more inequalities can be obtained from these.

The question arises then as to whether there exists a *complete* set of inequalities in the sense described in the introductory § 1 of this chapter. It is interesting to observe that this problem, which appears so *nonlinear* in form, can be transformed into a *linear* problem. In order to do this, we proceed in the following way. We first observe that s_k may be written as a Riemann-Stieltjes integral,

$$s_k = \int_0^\infty x^k\, dG(x), \tag{5}$$

where $G(x)$ is a step function with a jump of 1 at the points $x = x_i$, $i = 1, 2, \ldots, n$, assumed to be distinct and taken, with no loss of generality, to be monotone increasing, $0 \leqq x_1 < x_2 < \cdots < x_n$. If the x_i are constrained to lie between 0 and 1, the upper limit in (5) will be 1. The general case, where the x_i are not required to be distinct, can be subsumed under the problem of studying the power sums

$$t_k = \lambda_1 x_1^k + \lambda_2 x_2^k + \cdots + \lambda_n x_n^k, \tag{6}$$

where $\lambda_i > 0$ and the x_i are distinct.

Let us begin, then, by investigating the relationships that exist among the elements of the sequence $\{m_k\}$ determined by the integrals

$$m_k = \int_0^1 x^k\, dG(x), \tag{7}$$

where $G(x)$ is a monotone increasing bounded function over [0,1]. We shall call this sequence a *moment sequence,* and shall say that the m_k constitute the moments of the distribution function $dG(x)$.

This problem, of great importance in probability theory (see SHOHAT-TAMARKIN [1.11] and MALLOWS [1.13] for a discussion of it), has far too many ramifications to be adequately treated within the confines of one chapter of this monograph. What we propose to do here is to consider a few basic aspects of the area of research, with particular emphasis on some of the versatile techniques that can be applied and on their relevance to inequalities. For a discussion of the original problem connected with the s_k defined in (1), and also for some recent results, see URSELL [1].

The fundamental and elegant idea that we shall exploit here is *convexity*, following MINKOWSKI, CARATHÉODORY, and others.

§ 3. Convexity

Let $(m_1, m_2, \ldots, m_n)$ be the first n moment of the function $G(x)$, and $(m_1', m_2', \ldots, m_n')$ the corresponding first n moments of the function $H(x)$, where the moments are now computed by means of the formula (2.7). The linearity of the integrals in (2.7), as functionals of $G(x)$ and $H(x)$, enables us to assert that the quantities $\lambda m_1 + (1-\lambda) m_1', \ldots, \lambda m_n + (1-\lambda) m_n'$ constitute the first n moments of the function $\lambda G(x) + (1-\lambda) H(x)$. Furthermore, if $0 \leqq \lambda \leqq 1$, $\lambda G(x) + (1-\lambda) H(x)$ is a monotone increasing bounded function of x in $[0,1]$ whenever $G(x)$ and $H(x)$ are functions of this sort.

It follows that if we compute the first n moments of all distribution functions $G(x)$ defined over $[0,1]$, where $G(x)$ is monotone increasing and bounded there, and regard these n quantities as coordinates of a point in n-dimensional Euclidean space, then the set of points obtained in this way as $G(x)$ ranges over all functions of this type is *convex*. By this we mean, as usual, that whenever $P = (m_1, m_2, \ldots, m_n)$ and $Q = (m_1', m_1', \ldots, m_n')$ belong to this set, then *all* points $\lambda P + (1-\lambda) Q$, $0 \leqq \lambda \leqq 1$, on the line segment joining P and Q are also members of this set.

Let us now observe that this is not an isolated property of the moments defined above, but a general property shared by large classes of sequences defined by linear functionals.

Let f be an element in a space S, and let $\{\phi_k\}$ be a sequence of elements in a dual space S', where S and S' possess the property that an inner product (x, y) may be defined for any $x \in S$ and any $y \in S'$. Consider then the generalized moments

$$m_k = (f, \phi_k), \quad k = 1, 2, \ldots \tag{1}$$

If the elements ϕ_k possess no special properties, and the set S likewise has no distinguishing features, it is not to be expected that any interesting properties of the sequence $\{m_k\}$ will be uncovered. Suppose, however, that S possesses the important property that $f \in S$ and $g \in S$ implies that

$$h(x) = \lambda f(x) + (1-\lambda)\, g(x)$$

is also an element in S for all λ satisfying the inequalities $0 \leqq \lambda \leqq 1$. If S possesses this property, we shall say that it is *convex*. See BONNESEN and FENCHEL ([8.1] in Chapter 1) and EGGLESTON [1].

Then, just as above for the ordinary moments, it follows that the points $(m_1, m_2, \ldots, m_k)$ swept out as f runs through all elements in S constitute, for each k, a convex set of points in k-dimensional Euclidean space.

Our treatment of moment problems will hinge upon this fundamental notion of convexity, reinforced at various points by the equally fundamental idea of positivity.

§ 4. Some Examples of Convex Spaces

Let us now present some important examples of convex spaces and associated moment sequences.

A. $\qquad f \in S \;\text{ if }\; \int_0^1 f^2(x)\,dx \leqq 1; \quad \phi_k(x) \in L^2(0,1)\,.$

The Cauchy-Schwarz inequality asserts the existence of the moments

$$m_k = \int_0^1 f\phi_k\,dx; \tag{1}$$

and the triangle inequality, MINKOWSKI's inequality for $p = 2$, yields the convexity of the set S.

In the most important cases in which these moments occur, they are the Fourier coefficients corresponding to an orthonormal sequence $\{\phi_k(x)\}$.

B. $\quad f \in S \;\text{ if }\; \int_0^1 |f(x)|^p\,dx \leqq 1; \phi_k(x) \in L^q(0,1)\; p > 1,\; \frac{1}{p} + \frac{1}{q} = 1.$

The Hölder inequality asserts the existence of the moments, and the Minkowski inequality yields the convexity of S.

C. $\qquad f \in S \;\text{ if }\; \operatorname*{ess.\,supremum}_{0 \leqq x \leqq 1} |f(x)| \leqq 1; \phi_k(x) \in L\,(0,1)\,.$

D. $\qquad f \in S \;\text{ if }\; df \geqq 0\,, \int_0^1 df = 1; \phi_k(x) = e^{2\pi i k x}.$

This furnishes an example of a complex moment sequence.

§ 5. Examples of Nonconvex Spaces

Two interesting examples of nonconvex sets are the following:

A. $f(x) \in S$ if $f(x)$ assumes only two values, 0 or 1, for $0 \leqq x \leqq 1$; $\phi_k(x) = x^k$.

B. $f(x) \in S$ if $f(x)$ is a continuous unimodal function in $[0,1]$, i. e., a function possessing a single extremum in $[0,1]$; $\phi_k(x) = x^k$.

Both of these spaces can be made, by means of various artifices, to depend on convex spaces. We shall discuss one briefly (Bellman and Blackwell [1]), and shall refer the reader to the comprehensive paper by Mallows [1.13] for the other and for further references; see also Royden [2].

§ 6. On the Determination of Convex Sets

As we have stated in the foregoing sections, our aim is to present a method that can be used to generate a complete set of inequalities. To this end, we shall study the problem of determining the set of points produced by the set of moments $(m_1, m_2, \ldots, m_n)$, considered as a point in n-dimensional space, as f runs through the elements of a space S.

To determine this region, we seek to specify the boundaries of the region. In order to do this, we must exploit certain intrinsic properties of S. In particular, if S is convex, which implies that the set of points generated by the moments is convex, the following simple geometrical idea will play an essential role.

To locate the boundary points of a convex region R in n-dimensional space, take a plane P and move it parallel to itself until it contains extreme points of R. The set of extreme points obtained as we apply this procedure to all planes constitutes the boundary of R. This intuitive characterization is not always precise. For a rigorous and detailed discussion of this property, and for many further applications, see the book by Bonnesen and Fenchel ([8.1] in Chapter 1) and the papers by Rosenbloom [1.3].

As an example, consider a two-dimensional case. The lines L_e and L_E represent extreme positions of the line L, with P and P' determined as points belonging to the boundary B of the convex region R shown in Fig. 3 on the next page.

In general, each point on the boundary of R is furnished by a function $f \in S$ that yields a "tangent" plane. A tangent plane, in turn, is characterized very simply by the fact that its distance from the origin is a maximum or a minimum within the set of distances to the members of the family of parallel planes containing elements of the region R.

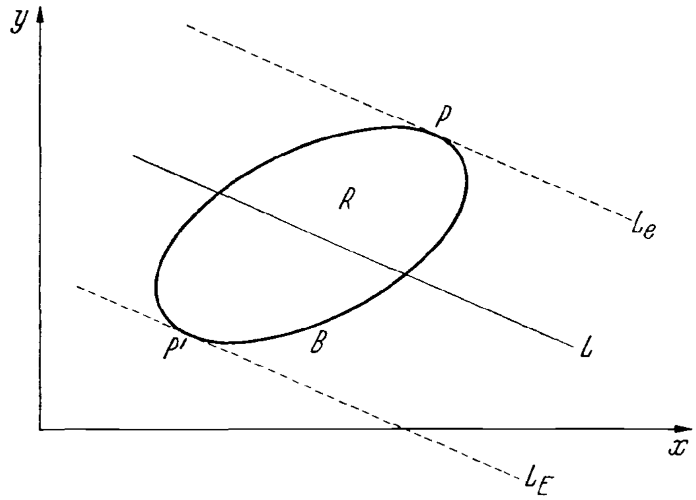

Fig. 3

Difficulties arise as far as this intuitive description is concerned when the boundary region has sharp corners and flat parts.

§ 7. L^p-Space — A Result of F. Riesz

As a first example of the technique introduced in § 6, consider the set S consisting of real functions $f(x)$ satisfying the constraint

$$\int_0^1 |f|^p \, dx \leqq 1 \,, \tag{1}$$

for a given $p > 1$.

Let $\{\phi_k(x)\}$ be a sequence of functions in L^q [0,1], and consider the moments determined by the integrals

$$m_k = \int_0^1 f \phi_k \, dx, \quad k = 1, 2, \ldots \tag{2}$$

Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the direction numbers of a family of parallel planes

$$\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_n x_n = r \,, \tag{3}$$

as r varies between $-\infty$ and $+\infty$. Let $(m_1, m_2, \ldots, m_n)$ be the coordinates of a point in the n-dimensional moment space determined by (2). Then for some $f(x)$ satisfying (1) we have

$$\int_0^1 f(x) \left[\sum_{k=1}^{n} \lambda_k \phi_k(x) \right] dx = r \,. \tag{4}$$

It remains to determine the elements in S that maximize or minimize r, a quantity directly proportional to the distance of the plane (3) from the origin. Using HÖLDER's inequality, we see that the extremal distances are furnished by the functions

$$f(x) = \frac{\left|\sum_{k=1}^{n} \lambda_k \phi_k\right|^{1/(p-1)} \operatorname{sgn}\left(\sum_{k=1}^{n} \lambda_k \phi_k\right)}{\left(\int_0^1 \left|\sum_{k=1}^{n} \lambda_k \phi_k\right|^q dx\right)^{1/q}}, \tag{5}$$

with

$$r_{\max} = \left(\int_0^1 \left|\sum_{k=1}^{n} \lambda_k \phi_k\right|^q dx\right)^{1/q}, \tag{6}$$

$$r_{\min} = -r_{\max}.$$

In order to make use of this result, we observe that a point is in the set R if and only if any plane through the point is at a distance from the origin that is between the $r_{\min}$ and $r_{\max}$ computed for the family of parallel planes determined by this plane.

We have thus established the following result; see F. RIESZ [1], [3], HELLY [2], and BANACH [1.1].

Theorem 1. *The point $(m_1, m_2, \ldots, m_n)$ is an element of the n-dimensional moment space determined as in* (1) *and* (2) *if and only if*

$$\left|\sum_{k=1}^{n} \lambda_k m_k\right| \leqq \left(\int_0^1 \left|\sum_{k=1}^{n} \lambda_k \phi_k\right|^q dx\right)^{1/q} \tag{7}$$

for every set of real numbers $(\lambda_1, \lambda_2, \ldots, \lambda_n)$.

Although this is an interesting and elegant result, it is only rarely applicable for $p \neq 2$; but see BOAS [4].

If, for a given infinite sequence $\{m_k\}$, (7) is satisfied for all n, then we can assert the existence of a function $f \in L^p\,[0,1]$, satisfying (1), for which (2) is valid. To see this, let $\{f_n\}$ be a sequence of functions such that

$$m_k = \int_0^1 f_n \phi_k \, dx, \quad k = 1, 2, \ldots, n, \tag{8}$$

with

$$\int_0^1 |f_n|^p \, dx \leq 1,$$

for $n = 1, 2, \ldots$ The sequence $\{f_n\}$ then possesses a subsequence $\{f_m\}$ weakly converging to a limit $f(x)$ for which

$$m_k = \lim_{n \to \infty} \int_0^1 f_n \phi_k \, dx = \int_0^1 f \phi_k \, dx. \tag{9}$$

§ 8. Bounded Variation

Let us now consider the situation (which turns out to be more fruitful analytically) in which $f(x)$ is a member of the set of monotone increasing functions over [0,1] having total variation 1, and in which $\{\phi_k(x)\}$ is a sequence of real continuous functions over this same interval.

The moment sequence is now determined by the Riemann-Stieltjes integrals

$$m_k = \int_0^1 \phi_k(x)\, df(x)\,, \quad k = 1, 2, \ldots \tag{1}$$

To determine the boundary of the convex set in n-dimensional space determined by $(m_1, m_2, \ldots, m_n)$, we shall determine the maximum and minimum of the expression

$$r = \int_0^1 \left[\sum_{k=1}^{n} \lambda_k \phi_k(x) \right] df(x) \tag{2}$$

as $f(x)$ varies over all functions $f(x)$, monotone increasing and bounded in [0,1].

It is clear that

$$\begin{aligned} r_{\max} &= \max_{0 \leqq x \leqq 1} \left[\sum_{k=1}^{n} \lambda_k \phi_k(x) \right], \\ r_{\min} &= \min_{0 \leqq x \leqq 1} \left[\sum_{k=1}^{n} \lambda_k \phi_k(x) \right]. \end{aligned} \tag{3}$$

In each case, the extremum is furnished by a function $f(x)$ that possesses a unit jump at a maximum or minimum, respectively, of the function $\sum_{k=1}^{n} \lambda_k \phi_k(x)$, and is constant elsewhere. If several maxima or minima exist, we obtain families of step functions that yield the extremum values.

We have thus derived the following result.

Theorem 2. *A necessary and sufficient condition that the points* $(m_1, m_2, \ldots, m_n)$ *belong to the n-dimensional moment space determined as in* (1) *is that*

$$\min_{0 \leqq x \leqq 1} \sum_{k=1}^{n} \lambda_k \phi_k(x) \leqq \sum_{k=1}^{n} \lambda_k m_k \leqq \max_{0 \leqq x \leqq 1} \sum_{k=1}^{n} \lambda_k \phi_k(x) \tag{4}$$

for all real sets of parameters $(\lambda_1, \lambda_2, \ldots, \lambda_n)$.

A precise determination of the boundary by means of the foregoing result depends on a knowledge of the possible number of maxima or minima that a function of the form

$$g(x) = \sum_{k=1}^{n} \lambda_k \phi_k(x) \tag{5}$$

can possess. Problems of this nature arise in the theory of best approximation initiated by ČEBYSHEV; see S. BERNSTEIN [1] and J. L. WALSH [2]. A detailed investigation of these questions for the case in which $\phi_k(x) = x^k$ is given in KARLIN and SHAPLEY [1.12], where other references may be found. An interesting investigation of the connection between these matters and the classical inequalities may be found in DRESHER ([24.1] in Chapter 1).

Questions of this type are also of interest in the determination of optimal strategies in the theory of games; see DRESHER and KARLIN [3] and KARLIN ([38.15] in Chapter 2).

§ 9. Positivity

Pursuing a different path, which will enable us to interlink these questions with the theory of positive definite quadratic forms, let us establish a variant of Theorem 2, namely, the following result:

Theorem 3. *A necessary and sufficient condition that $(m_1, m_2, \ldots, m_n)$ be a point in the n-dimensional moment space determined as in* (8.1) *is that*

$$\lambda_0 + \sum_{k=1}^{n} \lambda_k m_k \geqq 0 \tag{1}$$

for all values of $\lambda_0, \lambda_1, \ldots, \lambda_n$ such that

$$\lambda_0 + \sum_{k=1}^{n} \lambda_k \phi_k(x) \geqq 0 \tag{2}$$

for all x in [0,1].

Proof. It is clear that the condition is necessary. Let us then show that it is also sufficient. To do this, we show that (8.4) is implied by the conditions (1) and (2).

Suppose that (2) implies (1) and that there exists a set of parameters $(\lambda_1', \lambda_2', \ldots, \lambda_n')$ such that

$$\sum_{k=1}^{n} \lambda_k' m_k > \max_{0 \leqq x \leqq 1} \sum_{k=1}^{n} \lambda_k' \phi_k(x) . \tag{3}$$

Add a quantity λ_0' determined by the condition that

$$\lambda_0' + \max_{0 \leqq x \leqq 1} \sum_{k=1}^{n} \lambda_k' \phi_k(x) = 0 . \tag{4}$$

Then we have a sequence of values $(\lambda_0', \lambda_1', \ldots, \lambda_n')$ possessing the property that

$$\lambda_0' + \sum_{k=1}^{n} \lambda_k' m_k > 0 , \tag{5}$$

but

$$\lambda_0' + \sum_{k=1}^{n} \lambda_k' \phi_k(x) \leqq 0 \tag{6}$$

for all x in [0,1].

This contradicts the fact that (2) implies (1), since an obvious consequence of this condition is that

$$\lambda_0 + \sum_{k=1}^{n} \lambda_k m_k \leqq 0 \tag{7}$$

for all values of $\lambda_0, \lambda_1, \ldots, \lambda_n$ for which

$$\lambda_0 + \sum_{k=1}^{n} \lambda_k \phi_k(x) \leqq 0 \,. \tag{8}$$

The proof that our condition implies that

$$\min_{0 \leqq x \leqq 1} \sum_{k=1}^{n} \lambda'_k \phi_k(x) \leqq \sum_{k=1}^{n} \lambda'_k m_k \tag{9}$$

is precisely the same.

It is interesting to note that whereas we have used convexity arguments, this result may be readily derived from the Hahn-Banach theorem, as in Shohat-Tamarkin [1.11]; see also Banach [1.1]. This interchange of convexity and projection arguments with the fundamental theorem concerning the extension of linear functionals will be observed again in what follows; see Lax [1.19], Lorch ([25.5] of Chapter 1), Weston [1], and Ky Fan [1.15], each paper dealing with a different area of analysis.

The proof that the validity of these conditions for all n ensures the existence of a function $f(x)$, monotone increasing and bounded, follows the same lines as in § 7 if we replace the weak convergence theorem of L^p with the Helly convergence theorem for functions of bounded variation.

§ 10. Representation as Squares

To apply the criterion of Theorem 3, we must determine some necessary and sufficient conditions that a function of the form

$$g(x) = \lambda_0 + \lambda_1 \phi_1(x) + \cdots + \lambda_n \phi_n(x) \tag{1}$$

be nonnegative in [0,1]. Although no simple criteria exist for general sequences $\{\phi_k(x)\}$, very elegant criteria exist for the two most important cases,

$$\phi_k(x) = \cos 2\pi k x \,, \; \phi_k(x) = x^k, \quad k = 1, 2, \ldots \tag{2}$$

The general problem is itself, apart from any applications, one of great intrinsic interest, and one that lies at the very heart of the theory of inequalities.[1] Since the fundamental inequality, and one from which all others are deduced, states that a square of a real quantity is nonnegative, the problem naturally arises concerning the representation of a particular nonnegative quantity as a square. The problem in this general

[1] Cf. our discussion in §§ 45 and 47 of Chapter 1 and § 4 of Chapter 2.

form is trivial, i. e.,

$$x = (\sqrt{x})^2, \tag{3}$$

for $x \geqq 0$. More subtle questions arise, however, when we turn to the possibility of this representation under the further condition that $\sqrt{x}$ be an element of a certain preferred set, such as the set of polynomials, polynomials with real coefficients, rational functions, trigonometric polynomials, and so on.

Barring this, we may ask whether a nonnegative element can be written as a sum of squares,

$$x = x_1^2 + x_2^2 + \cdots. \tag{4}$$

We have already met an example of this representation in §§ 3–5 of Chapter 2 on quadratic forms.

The general problem occurs in many fields of mathematics — e. g., WARING's problem in analytic number theory, the Hilbert-Artin problem in algebra, and the representation of nonnegative harmonic functions in analysis (see the discussions given in the book by SZEGÖ [1]). Here we shall merely cite two specific results, one pertaining to the sequence $\{\cos 2k\pi x\}$, and one to the sequence $\{x^k\}$. Proofs of these results, and a number of further results, may be found in PÓLYA and SZEGÖ ([44.1] of Chapter 1); see also KARLIN and SHAPLEY [1.12], and SHOHAT and TAMARKIN [1.11].

§ 11. Nonnegative Trigonometric and Rational Polynomials

The first result we shall require for the representation of nonnegative polynomials is due to F. RIESZ:

Theorem 4. *If $g_n(x)$ is a nonnegative cosine polynomial,*

$$g_n(x) = \lambda_0 + \lambda_1 \cos 2\pi x + \cdots + \lambda_n \cos 2n\pi x, \tag{1}$$

for $0 \leqq x \leqq 1$, then

$$g_n(x) = |x_0 + x_1 e^{2\pi i x} + \cdots + x_n e^{2\pi i n x}|^2, \tag{2}$$

where the x_i are real.

The second result, which is a consequence of Theorem 4, is the following:

Theorem 5. *If $g_n(x)$ is a rational polynomial,*

$$g_n(x) = \lambda_0 + \lambda_1 x + \cdots + \lambda_n x^n, \tag{3}$$

that is nonnegative in $-1 \leqq x \leqq 1$, then

$$g_n(x) = [p(x)]^2 + (1 - x^2)\,[q(x)]^2, \tag{4}$$

where $p(x)$ and $q(x)$ are polynomials of degree n or $n-1$ with real coefficients.

For proofs, the reader may consult Pólya and Szegö ([44.1] of Chapter 1). These results, together with the results of the foregoing sections, can be used to obtain a number of necessary and sufficient conditions in the theory of moments.

§ 12. Positive Definite Quadratic Forms and Moment Sequences

A fundamental result is the following theorem of Carathéodory:

Theorem 6. *A necessary and sufficient condition that the sequence* $\{m_k\}$ *be representable as a sequence of trigonometric moments,*

$$m_k = \int_0^1 \cos 2\pi\, k\, x\, df(x)\,, \tag{1}$$

where $df \geqq 0$, *is that the quadratic forms*

$$\sum_{k=0}^{n} m_k \left(\sum_{|i-j|=k} x_i x_j \right) \tag{2}$$

be nonnegative, or, equivalently, that we have the determinantal inequalities

$$|m_{|i-j|}| > 0\,, \quad i, j = 0, 1, \ldots, n; \quad n = 1, 2, \ldots \tag{3}$$

The necessity is obvious, as we see upon forming the quadratic form

$$\int_0^1 \left| \sum_{k=0}^{n} x_k\, e^{2\pi i k x} \right|^2 df(x)\,. \tag{4}$$

The sufficiency requires more.

There are a number of interesting ways of establishing this result of Carathéodory. A direct proof based on the canonical representation of positive definite quadratic forms may be found in Fischer [1]. It is reproduced in [2.1] of Chapter 2, where a number of further references to works of Szász, Fejér, Schur, and Carathéodory are given. For other proofs, see the books by Ky Fan [1.7] and Shohat-Tamarkin [1.11].

§ 13. Historical Note

In a brief paper that gave little indication of the important role it would play in the development of mathematical analysis, Picard [1] established in a very simple fashion his famous theorem concerning the set of values that may be omitted by an entire function. Both the elegance of the result and the depth of the method used to prove it gave rise to a series of interesting and significant investigations devoted to the understanding and extension of theorems of this nature; see Borel [2], Landau [3], [9], and Bieberbach [8].

In the course of sharpening Picard's theorem, Landau was led to the study of harmonic functions that are nonnegative inside the unit circle. Following the type of investigation made classic by the thesis of HADAMARD, the question arose as to whether this property could be detected by a knowledge of the Fourier coefficients of the function. This problem was studied intensively by FEJÉR [4], CARATHÉODORY [5], CARATHÉODORY and FEJÉR [6], HERGLOTZ [7], FISCHER [12.1], and F. RIESZ [7.3].

It was recognized by CARATHÉODORY that this was a moment problem. Introducing the techniques of convexity that we have used in the foregoing §§ 3–12, he furnished a rigorous foundation for these methods.

Closely related to the researches mentioned above is the study initiated by TOEPLITZ. Let $f(\theta)$ be a real function of θ defined over $0 \leqq \theta \leqq 2\pi$, and let $\{c_n\}$, $n = 0, \pm 1, \pm 2, \ldots$, be the sequence of Fourier coefficients

$$c_n = \frac{1}{2\pi} \int_0^{2\pi} f(\theta)\, e^{-in\theta}\, d\theta\,. \tag{1}$$

The finite matrices $T_n = (c_{k-l})$, $k, l = 0, 1, \ldots, n-1$, are called *Toeplitz matrices*. The characteristic roots of these matrices are closely related to the values assumed by $f(\theta)$ in the interval $[0, 2\pi]$; see G. PÓLYA and G. SZEGÖ ([44.1] of Chapter 1).

The study of these matrices has been resumed in recent years because of their important role both in probability theory and in statistical mechanics in the study of certain idealized "order-disorder" problems; see the papers by KAC, MURDOCK, and SZEGÖ [10], and MARADUDIN and WEISS [12], and the book by GRENANDER and SZEGÖ [11].

§ 14. Positive Definite Sequences

It is clear from what has preceded that a concept important in its own right is that of a positive definite sequence:

A sequence of complex numbers $\{a_n\}$, $n = 0, \pm 1, \pm 2, \ldots$, is called *positive definite* if

$$\begin{aligned} &\text{(a)} \qquad a_{-n} = \overline{a_n}\,, \\ &\text{(b)} \qquad \sum_{k,j=1}^{n} a_{k-j}\, z_k \bar{z}_j \geqq 0\,, \end{aligned} \tag{1}$$

for all sets of complex z_k.

The fundamental representation theorem is due to F. RIESZ [7.1]:

Theorem 7. *A necessary and sufficient condition that a sequence $\{a_n\}$ be positive definite is that*

$$a_n = \int_0^{2\pi} e^{in\theta}\, dV(\theta)\,, \tag{2}$$

where $V(\theta)$ is a real monotone nondecreasing function of bounded variation.

It is interesting to note that HERGLOTZ's paper [13.7] is also devoted to the Picard-Landau-Carathéodory problem area. Furthermore, a derivation of this result using only the representation theory of positive definite quadratic forms, plus the Helly selection theorem, may be found in the paper by FISCHER [12.1] to which we have previously referred; a more accessible reference is BELLMAN ([2.1] in Chapter 2).

An excellent demonstration of the central role that this representation plays in the field of modern analysis is contained in the paper by KY FAN [1], where many references are given to its use in probability theory; see also the book by GRENANDER and SZEGÖ [13.11].

§ 15. Positive Definite Functions

A natural and most important extension of the concept introduced in the foregoing section is that of a positive definite function. As defined by MATHIAS [1.4], a complex function defined for all real x is said to be positive definite if it satisfies the following conditions:

(a) $f(x)$ is bounded and continuous for all x, $-\infty < x < \infty$.

(b) $f(-x) = \overline{f(x)}$.

(c) For any set of real values $x_1, x_2, \ldots, x_n$ and any complex values $c_1, c_2, \ldots, c_n$, we have

$$\sum_{r,s=1}^{n} c_r \overline{c_s} f(x_r - x_s) \geqq 0 .$$

The fundamental representation theorem is due to BOCHNER; see FAN [1.5].

Theorem 8. *Under the above conditions (a), (b), (c), there exists a nondecreasing bounded function $g(y)$ such that*

$$f(x) = \int_{-\infty}^{\infty} e^{iyx} dg(y)$$

for all real x.

This result occupies a basic position in the modern theory of harmonic analysis. It has been extensively generalized; see COOPER [1.6], LOOMIS [1], WEIL [2], SCHWARTZ [1.22], and the papers by GODEMONT [1.9], CARTAN and GODEMONT [1.10], GELFAND, and RAIKOV referred to in the introductory section of this chapter. More recent results may be found in DEVINATZ [3].

Generalized positive definite functions arise in connection with the problem of determining when a scalar complex function, defined for elements of a topological space, is a Hilbert distance function. See BOCHNER [4], SCHOENBERG [5], and SCHOENBERG and VON NEUMANN [6] for a discussion of these matters.

Positive definite functions of generalized type also play a role in the Pick-Nevanlinna interpolation theory; see Sz.-Nagy and Korányi [7]. Since these problems can also be considered to be moment problems, (see Weyl [8]), we once again see an interweaving of different strands of analysis.

§ 16. Reproducing Kernels

A complex function $k(x, y)$ is called a *reproducing kernel* if

$$k(x, y) = \int_{-\infty}^{\infty} k(x, t)\, \overline{k(t, y)}\, dt \tag{1}$$

for all real x and y. These kernels enter analysis by way of the study of Green's functions; see Aronszajn and Smith [2].

The basic representation theorem is due to Aronszajn [1]:

Theorem 9. *A complex function $k(x, y)$ is a reproducing kernel if and only if*

$$\sum_{i,j=1}^{n} c_i \overline{c_j}\, k(x_i, x_j) \geqq 0 \tag{2}$$

for all complex c_i and all real x_i.

§ 17. Nonconvex *Spaces*

It can happen that the n-dimensional moment region is convex, even though the underlying space S is not. An interesting example of this phenomenon is the first example discussed in § 5. There we consider the space S of all characteristic functions $f(x)$ defined as follows:

$$\begin{aligned} f(x) &= 1 \text{ if } x \in E, \text{ where } E \text{ is a given subset of } [0,1], \\ &= 0 \text{ otherwise,} \end{aligned} \tag{1}$$

and the sequence of functions is given by $\phi_k(x) = x^k$. The problem we set ourselves is that of determining the region swept out by the points

$$\left(\int_E dx, \int_E x\,dx, \ldots, \int_E x^n\,dx\right), \tag{2}$$

as E ranges over all Borel subsets of $[0,1]$.

A theorem of A. Liapounoff [1] guarantees that this region is closed and convex, despite the fact that the space of characteristic functions $f(x)$ clearly is *not* convex. Using this fact, we can easily derive (Bellman and Blackwell [5.1]) the following result.

Theorem 10. *Let X_j denote the interval $\{a_j \leqq x \leqq a_{j+1}\}$ for $j = 0, \ldots, n$, where $a_0 = 0 \leqq a_1 \leqq a_2 \leqq \cdots \leqq a_{n+1} = 1$. The sets E_0, E_1, defined by*

$$\begin{aligned} E_0 &= X_0 \cup X_2 \cup \cdots, \\ E_1 &= X_1 \cup X_3 \cup \cdots, \end{aligned} \tag{3}$$

yield boundary points of the moment space, and every E yielding a boundary point is, except for sets of measure zero, of one of these forms.

It follows that a parametric representation of the boundary — which consists of two parts, D_1 and D_2 — is given by

$$\begin{aligned} D_1&: \ x_i = r_{i0}(a) + r_{i2}(a) + \cdots, \\ D_2&: \ x_i = r_{i1}(a) + r_{i3}(a) + \cdots, \quad i = 0, 1, \ldots, n, \end{aligned} \tag{4}$$

where

$$r_{ij}(a) = \frac{a_{j+1}^{i+1} - a_j^{i+1}}{i+1}, \quad i, j = 0, 1 \ldots, n. \tag{5}$$

Proof. The functional

$$r = \int\limits_E \left(\sum_{i=0}^{n} \lambda_i t^i\right) dt \tag{6}$$

is minimized over all E for the set

$$E = \left(t, \sum_{i=0}^{n} \lambda_i t^i \leqq 0\right). \tag{7}$$

This set is of form either E_0 or E_1 for suitably chosen a_i, and conversely every E_0 or E_1 furnishes a boundary set for suitably chosen λ_i. Observe that in this case the parametric representation of boundary elements is obtained most easily in coordinates that are quite different from the λ_i.

The same type of argument yields the region swept out by

$$\left(\int\limits_{E_0} dx, \int\limits_{E_1} x\, dx, \ldots, \int\limits_{E_n} x^n\, dx\right), \tag{8}$$

as $(E_0, E_1, \ldots, E_n)$ varies over all partitions of $0 \leqq x \leqq 1$ into $n+1$ disjoint Borel sets $E_0, E_1, \ldots, E_n$. Details may be found in [**5.1**].

The second example mentioned in § 5, continuous functions over [0,1] that possess a unique extremum, may be treated in the above fashion by first considering the subspace of functions possessing a unique maximum at a fixed point $x = a$, and then taking the envelope over a. For a thorough exposition of moment problems in probability theory, see the previously cited paper by MALLOWS [1.13] and ROYDEN [5.2].

§ 18. A "Resonance" Theorem of Landau

Let us now consider some "interaction" theorems of a different type, beginning with a convergence result due to LANDAU [1]. As in the case of the Picard theorem, the result seems quite special, and does not in any way presage its many significant extensions.

Theorem 11. *A necessary and sufficient condition that the series*

$$\sum_{n=1}^{\infty} |a_n|^p, \quad p > 1,$$

converge is that the series

$$\sum_{n=1}^{\infty} a_n b_n$$

converge for all sequences $\{b_n\}$ *for which*

$$\sum_{n=1}^{\infty} |b_n|^q$$

converges, $q = p/(p-1)$.

The necessity we recognize as an immediate consequence of the Hölder inequality. The sufficiency is the meat of the result. It is not difficult to give a proof by means of contradiction, using an interesting result due to Abel concerning the convergence and divergence of series of the form

$$\sum_{n=1}^{\infty} \frac{u_n}{s_n^a}, \quad 0 < a < \infty , \tag{1}$$

where $s_n = u_1 + u_2 + \cdots + u_n$. As we shall see in § 19, however, the result is actually a special case of a very powerful theorem of analysis, and any local proof only obscures the significance of the result. See BANACH [1.1] for a proof using the Banach-Steinhaus theorem.

Along the same lines, we mention a generalization of classical results concerning Cauchy products.

Theorem 12. *A necessary and sufficient condition that the series* $\sum_{n=1}^{\infty} c_n$, *where*

$$c_n = \sum_{k=0}^{n} a_k b_{n-k} , \tag{2}$$

converge for all convergent series $\sum_{n=0}^{\infty} b_n$ *is that*

$$\sum_{n=0}^{\infty} |a_n| < \infty .$$

There are extensive generalizations to various modes of summability. All these are particular applications of the Hurwitz-Silverman-Toeplitz summability theorem (see ZYGMUND [1.2]), which is, in turn, a particular application of the theorem we shall discuss immediately below.

Results of this nature have been discovered and rediscovered by so many mathematicians, in published and unpublished form, that it is difficult to assign authorship.

§ 19. The Banach-Steinhaus Theorem

Consider the Banach space $l^p, p > 1$, consisting of all sequences $\{a_n\}$ for which the series $\sum_{n=1}^{\infty} |a_n|^p$ satisfies

$$\sum_{n=1}^{\infty} |a_n|^p < \infty .$$

For an infinite sequence $b = \{b_n\}$, consider the linear functional defined by the inner product

$$f(a) = (a, b) = \sum_{n=1}^{\infty} a_n b_n . \tag{1}$$

Theorem 11 asserts that the finiteness of $f(a)$ for *all* elements $b \in l^q$ ensures the uniform boundedness of $f(a)$ for all b in the sphere $||b|| \leqq 1$, where

$$||b|| = \left(\sum_{n=1}^{\infty} |b_n|^q\right)^{1/q} . \tag{2}$$

This is equivalent to the statement that

$$||a|| = \left(\sum_{n=1}^{\infty} |a_n|^p\right)^{1/p}$$

is finite.

Stated in these terms, the result of LANDAU is a forerunner, and particular example, of the Banach-Steinhaus theorem concerning linear functionals; see BANACH [1.1] and ZYGMUND [1.2].

This result occupies a fundamental place in the theory of orthogonal series, in connection with multipliers and Lebesgue constants, in the theory of summability, as mentioned above, and even enters into the theory of differential equations. BELLMAN [1], and MASSERA and SCHÄFFER [2], [3], characterized the solutions of the homogeneous vector-matrix equation, or more generally, an equation in Banach space,

$$\frac{dx}{dt} = A(t)\, x , \tag{3}$$

in terms of the solutions of *all* equations of the form

$$\frac{dy}{dt} = A(t)\, y + f(t) . \tag{4}$$

The original results in this field were obtained by PERRON. See also CESARI [4] and CORDUNEANU [5]. Further references will be found in the papers cited above.

§ 20. A Theorem of Minkowski

Directly in line with the foregoing ideas, but with a quite different measure of interaction, is the following elegant result of MINKOWSKI.

Theorem 13. *Let x and y be n-dimensional vectors, and let A be an m by n matrix. Then a necessary and sufficient condition that $(x, y) \geqq 0$ for all y such that $Ay \geqq 0$ is that $x = A'b$, where b is an m-dimensional vector satisfying $b \geqq 0$.*

The notation $b \geqq 0$ for vectors means that each component of b is nonnegative.

Proofs of this result may be found in the recent book of D. GALE [1], and H. WEYL [2]. See FAN [1.15] for detailed discussions of this field.

It is interesting to note that the arguments concerning convex sets, which play such a dominant role in these papers, were, as indicated above, first developed by CARATHÉODORY in his paper devoted to the trigonometric moment problem, while the support function of MINKOWSKI, used by WEYL in his paper, is precisely the function utilized for quasi linearization in previous chapters. As we repeatedly see, a few basic techniques, themselves closely intertwined, dominate the field of inequalities and, indeed, much of classical and modern analysis.

It is also interesting to note the repeated parallelism between the theory of Hermitian operators and the theory of positive operators.

§ 21. The Theory of Linear Inequalities

The theorem of MINKOWSKI in § 20 is a particular result in the general theory of linear inequalities, much of which centers about the problem of maximizing a linear form

$$L(x) = \sum_{i=1}^{n} c_i x_i \,, \tag{1}$$

over all x_i satisfying a set of linear inequalities

$$\sum_{j=1}^{n} a_{ij} x_j \leqq b_i \,, \quad i = 1, 2, \ldots, m \,. \tag{2}$$

An enormous amount of work has been done in this field in recent years. Some particularly noteworthy papers ranging over the years are MINKOWSKI [1], FARKAS [2], STIEMKE [3], DINES [4], CARVER [5], MOTZKIN [6], ČERNIKOV [7], FAN [1.15], and NIKAIDÔ [12].

In recent years, the subject has again become a focus of attention as a consequence of its many applications in applied mathematics, particularly in the theory of games.

The emphasis has shifted to the problem of determining algorithms that yield numerical solutions of the maximization problem described above, with particular reference to operations that quickly and accurately can be performed by means of digital computers. The principal technique at the moment is the "simplex method" of G. DANTZIG [8], together with its refinements and extensions by CHARNES, LEMKE, BEALE, and others. For problems of this general class with various types of structure, there exist special algorithms; see FORD and FULKERSON [9] and BELLMAN ([7.1] of Chapter 1).

The topic of obtaining algorithms yielding computational solutions of the foregoing minimization problem is called "the theory of linear programming." It occupies an important place in operations research and mathematical economics. See DANTZIG, ORDEN, and WOLFE [8], RILEY and GASS [10], KARLIN ([38.15] in Chapter 2), and DORFMAN, SAMUELSON, and SOLOW ([42.4] in Chapter 2).

For an interesting application of inequalities of this nature to the foundations of probability theory along the lines of "rational betting," see LEHMAN [11].

§ 22. Generalizations

There have been two types of generalizations of the theory referred to in § 21. The first goes in the direction of nonlinear functions (see GALE, KUHN, and TUCKER [1], and FAN, GLICKSBERG, and HOFFMAN [1.14]), while the second studies linear functionals in more general spaces (see FAN [1.15], DUFFIN [2], LEHMAN [3], and BELLMAN and LEHMAN [4]). In § 24, below, we shall discuss a significant extension of the latter type.

The work of FAN [1.15] is particularly interesting in view of what we have discussed in the foregoing §§ 18–21, in that his work brings together linear inequalities and approximation theory by way of the general theory of linear and Banach spaces. In particular, the Hahn-Banach extension theorem and convexity theorems play vital roles.

§ 23. The Min-max Theorem of von Neumann

As commented in § 21, there is a close connection between the theory of linear inequalities and the theory of games created by VON NEUMANN; see VON NEUMANN and MORGENSTERN [1], WILLIAMS [3], and earlier work by É. BOREL [2].

The cornerstone of the theory of games is a variational problem of deceptively simple appearance:

Theorem 14. *Let x denote the set of values satisfying the conditions*

$$\begin{aligned} &\text{(a)} \quad x_i \geqq 0\,, \\ &\text{(b)} \quad \sum_{i=1}^{n} x_i = 1\,, \end{aligned} \tag{1}$$

and let y denote a similar set. If (a_{ij}) is any matrix of real quantities, then

$$\min_y \max_x \sum_{i,j=1}^{n} a_{ij} x_i y_j = \max_x \min_y \sum_{i,j=1}^{n} a_{ij} x_i y_j\,. \tag{2}$$

Proofs of this result may be found in [1], [4]. The result may be derived in a number of ways from the theory of linear inequalities referred to in the foregoing §§ 21, 22; established by means of fixed-point theorems, a method that leads to many generalizations; established inductively (Loomis [4]); or derived variously as the "steady state" of a dynamic process (J. Robinson [5]). An interesting proof based on the laws of static mechanics is due to Gross [6].

§ 24. The Neyman-Pearson Lemma

The classical inequalities can be established by standard variational techniques because of the occurence of nonlinearities. These nonlinearities play essential roles in the derivation of the conditions that characterize an extremum.

A number of interesting questions in analysis and applications — in particular, the theory of control processes — give rise to problems in which we wish to minimize a linear functional of the form

$$\int_0^1 f(t)\, a(t)\, dt \tag{1}$$

over all functions $f(t)$ subject to the constraints

$$\begin{aligned} &\text{(a)} \quad 0 \leqq f(t) \leqq 1\,,\ 0 \leqq t \leqq 1\,, \\ &\text{(b)} \quad \int_0^1 f(t)\, b(t)\, dt \leqq c\,, \end{aligned} \tag{2}$$

where $a(t)$ and $b(t)$ are given functions, and c is a known constant.

As we know from the preceding discussion, this problem is equivalent to that of determining the region in two-dimensional space spanned by the moments

$$\begin{aligned} m_1 &= \int_0^1 f(t)\, a(t)\, dt\,, \\ m_2 &= \int_0^1 f(t)\, b(t)\, dt\,, \end{aligned} \tag{3}$$

as $f(t)$ ranges over the space S of functions satisfying the constraint (2a).

In order to furnish a sample of the results that are obtained in variational problems of this type, and the methods used, we shall give a detailed proof of the following result.

Theorem 15. *The solution to the foregoing variational problem, under the further conditions that*

$$\text{(a)}\quad b(t) \geqq 0\,, \tag{4}$$
$$\text{(b)}\quad c \geqq 0\,,$$

is determined as follows:

Let set functions $E^- = E^-(k)$, $E = E(k)$, $E^+ = E^+(k)$ be defined as follows for $-\infty < k < \infty$:

$$\begin{aligned} E^-(k) &= [t; a(t) < k\,b(t)]\,,\\ E(k) &= [t; a(t) = k\,b(t)]\,,\\ E^+(k) &= [t; a(t) > k\,b(t)]\,. \end{aligned} \tag{5}$$

Determine k_0 by the condition that k_0 be the supremum over all nonpositive k satisfying the inequality

$$\int_{E^-} b(t)\,dt \leqq c\,, \tag{6}$$

and let

$$E^-(k_0) = E_0^-,\ E(k_0) = E_0,\ E^+(k_0) = E_0^+\,. \tag{7}$$

Then the set of minimizing functions f^* is given by

$$\begin{aligned} &\text{(a)}\quad f^*(t) = 1 \text{ on } E_0^-,\\ &\text{(b)}\quad f^*(t) = 0 \text{ on } E_0^+\,,\\ &\text{(c)}\quad f^*(t) = \text{arbitrary on } E_0, \text{ satisfying only the conditions (2a) and } \int_0^1 f^*(t)\,b(t)\,dt = c \text{ if } k_0 < 0, \text{ or the conditions (2a) if } k_0 = 0. \end{aligned} \tag{8}$$

The solution to the corresponding maximization problem is determined similarly.

Proof. For any $f(t)$ satisfying (2), we write

$$\begin{aligned} \int_0^1 f(t)\,a(t)\,dt = \int_{E_0^-} + \int_{E_0} + \int_{E_0^+} &= \int_{E_0^+} f(t)\,a(t)\,dt\\ &+ k_0 \int_{E_0} f(t)\,b(t)\,dt + \int_{E_0^-} f(t)\,a(t)\,dt\,. \end{aligned} \tag{9}$$

Since

$$\begin{aligned} \int_0^1 f^*(t)\,a(t)\,dt &= \int_{E_0^-} + \int_{E_0}\\ &= \int_{E_0^-} a(t)\,dt + k_0 \int_{E_0} f^*(t)\,b(t)\,dt\\ &= \int_{E_0^-} a(t)\,dt + k_0 \int_{E_0} f^*(t)\,b(t)\,dt\\ &= \int_{E_0^-} a(t)\,dt + k_0 \left[c - \int_{E_0^-} b(t)\,dt\right], \end{aligned} \tag{10}$$

we have

$$\begin{aligned}\Delta &= \int_0^1 f(t)\, a(t)\, dt - \int_0^1 f^*(t)\, a(t)\, dt \\ &= \int_{E_0^-} [f(t) - 1]\, a(t)\, dt + k_0 \int_{E_0} f(t)\, b(t)\, dt \\ &+ \int_{E_0} f(t)\, a(t)\, dt - k_0 \left[c - \int_{E_0^-} b(t)\ dt\right].\end{aligned} \tag{11}$$

Using the fact that $a(t) > k_0\, b(t)$ on E_0^+, we obtain the continued inequality

$$\begin{aligned}\Delta &\geqq \int_{E_0^-} [f(t) - 1]\, a(t)\, dt \\ &+ k_0 \left[\int_{E_0} f(t)\, b(t)\, dt + \int_{E_0^+} f(t)\, b(t)\, dt\right] - k_0 \left[c - \int_{E_0^-} b(t)\ dt\right] \\ &\geqq \int_{E_0^-} [f(t) - 1]\, a(t)\, dt + k_0 \left[c - \int_{E_0^-} f(t)\ b(t)\ dt\right] \\ &- k_0 \left[c - \int_{E_0^-} b(t)\ dt\right] \\ &\geqq \int_{E_0^-} [f(t) - 1]\, a(t)\, dt + k_0 \int_{E_0^-} [1 - f(t)]\, b(t)\, dt \\ &\geqq \int_{E_0^-} [1 - f(t)]\, [k_0\, b(t) - a(t)]\, dt \geqq 0\,.\end{aligned} \tag{12}$$

Since the signs of equality hold throughout if and only if $f(t)$ is an $f^*(t)$, we see that the $f^*(t)$ constitute the totality of minimizing functions.

§ 25. Orthogonal Projection

In the preceding sections, we have examined a number of situations in which a function f has been characterized by means of moments, (f, ϕ_n), taken with respect to a sequence of functions $\{\phi_n\}$. These moments may be regarded as *projections* of f on the "axes" ϕ_n. The theory of orthogonal series (Zygmund [1.2] and Kaczmarz and Steinhaus [1]), vast as it is, is only a particular study of this type.

The idea can be extended considerably. One of the most fruitful extensions is to the theory of partial differential equations, starting with the concept of orthogonal projection due to Weyl [1.18]. Preliminary results in this direction were obtained by Zaremba; see Rosenbloom [1.3] and Forsythe and Rosenbloom [2].

The basic idea is the following. If u is a solution of the linear equation $L(u) = 0$, then clearly

$$(L(u), v) = 0 \tag{1}$$

for all v for which the inner product is defined. If, in addition, v permits the application of the operator adjoint to L, which we call M, then (1) leads to the result

$$(L(u), v) = (u, M(v)) = 0 . \tag{2}$$

Suppose then we begin with the relation

$$(u, M(v)) = 0 , \tag{3}$$

for *all* v in a suitably chosen set. Under favorable conditions, this forces u to be a solution of $L(u) = 0$.

It is interesting to note that a number of problems that have been treated by the methods of orthogonal projection can also be treated by means of the Hahn-Banach extension theorem; see LAX [1.19].

When we state the results of the previous type in the foregoing fashion, we see their abstract identity with various versions of what is often called the fundamental lemma of the calculus of variations; see COURANT and HILBERT, ([26.1] in Chapter 2) and BERWALD [1.21]. Here, the aim is to conclude that the Euler equation, $E(u) = 0$, is a consequence of relations of the form

$$(E(u), v) = 0 , \tag{1}$$

for all v in a prescribed set of functions.

The foregoing ideas lead naturally to the concept of generalized solutions of linear and nonlinear operator equations, a basic contribution of BOCHNER and FRIEDRICHS. The organization of these ideas leads to the theory of distributions of SCHWARTZ [1.22].

§ 26. Equivalence of Minimization and Maximization Processes

Intimately related to the duality we have constantly emphasized and exploited is the fact that a number of minimization problems can be shown to be equivalent to maximization problems. The importance of this identity for the derivation of upper and lower bounds is clear. A discussion of results of FRIEDRICHS in the calculus of variations will be found in COURANT and HILBERT ([26.1] in Chapter 2), and application and further references in LAX [1] and ROGOSINSKI and SHAPIRO [2].

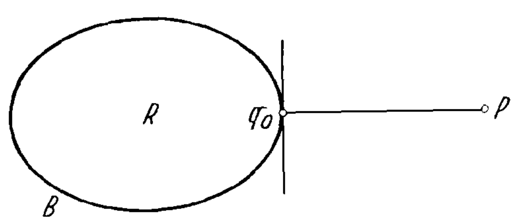

Fig. 4

A common origin of many of these results is the following simple geometrical property. Let R be a convex region with boundary B, and let p be a point outside R, as shown in Fig. 4. Consider the problem of minimizing the distance, $\|p - q\|$, between p and a point $q \in R$. The minimum will be attained for a point $q_0 \in B$, and the line pq_0 will be orthogonal to the tangent plane at q_0.

Now consider the set of tangent planes to B, and the distances from p to these planes; see Fig 5. It is clear that a local maximum for these distances will be furnished by the distance, $\|p - q_0\|$, to the plane tangent at q_0.

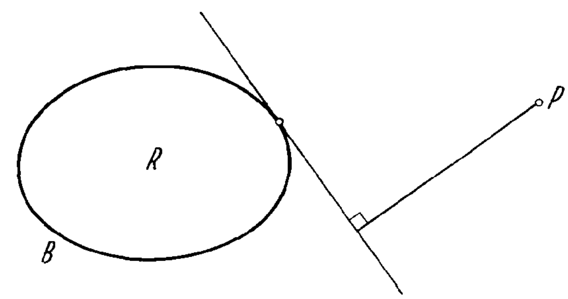

Fig. 5

We see then how the dual description of a surface, locus of points = envelope of tangents, naturally leads to the equivalence of certain minimization and maximization problems. The abstract-space version of this result leads to a number of interesting and important results.

Bibliographical Notes

§ 1. The fundamental reference for the theory of Banach spaces is

1. BANACH, S.: Théorie des operations linéaires. Monografje Matematyczne I. Warsaw: Z Subwencji Funduszu Kultury Narodowej 1932.

A number of results particularly applicable to the study of orthogonal series may be found in Chapter 4 of

2. ZYGMUND, A.: Trigonometrical series. Monografje Matematyczne V. Warsaw: Z Subwencju Funduszu Kultury Narodowej 1935.
3. ROSENBLOOM, P.: Quelque classes de problèmes extrémaux. Bull Soc. Math. France **79**, 1—58 (1951); **80**, 183—215 (1952).

Some fundamental references in the theory of positive definite functions and sequences are the following:

4. MATHIAS, M.: Über positive Fourier-Integrale. Math. Z. **16**, 103—125 (1923).
5. BOCHNER, S.: Vorlesungen über Fouriersche Integrale. Leipzig: Akademische Verlagsgesellschaft 1932.
6. COOPER, J. L. B.: Positive definite functions of a real variable. Proc. London Math. Soc. **10**, 53—66 (1960).
7. FAN, K.: Les fonctions définies-positives et les fonctions completement monotones. Mémorial Sci. Math. **64**. Paris: Gauthier-Villars 1950.
8. WIDDER, D. V.: The Laplace transform. Princeton, N. J.: Princeton University Press 1941.

Extensive generalizations of these concepts in abstract setting are given in

9. GODEMONT, R.: Les fonctions de type positif et la théorie des groupes. Trans. Am. Math. Soc. **63**, 1—84 (1948).
10. CARTAN, H., and R. GODEMONT: Théorie de la dualité et analyse harmonique dans les groupes abéliens localement compacts. Ann. Sci. École Norm. Sup. (3) **64**, 79—99 (1947).

Discussions of moment theory will be found in the work by WIDDER [8] cited above, and in

11. SHOHAT, J., and J. TAMARKIN: The problem of moments. Mathematical Surveys, vol. 2. New York: American Mathematical Society 1943.
12. KARLIN, S., and L. SHAPLEY: Geometry of moment spaces, Memoirs of the American Mathematical Society, 12. New York: American Mathematical Society 1953.

For a detailed discussion of many particular classes of moment problems, an excellent survey of the field, and the connection with probability and statistics, see

13. MALLOWS, C. L.: Generalizations of TCEBYCHEFF's inequalities. J. Roy. Stat. Soc. (B) **18**, 139—176 (1956).

We have systematically avoided any discussion of a large number of interesting and important inequalities arising from statistical sources. Any satisfactory discussion would materially increase the size of this monograph.

14. FAN, K., I. GLICKSBERG and A. HOFFMAN: Systems of inequalities involving convex functions. Proc. Am. Math. Soc. **8**, 617—622 (1957).
15. FAN, K.: On systems of linear inequalities. Ann. Math. Stud. **38**, 99—156. Princeton, N. J.: Princeton University Press 1956.
16. FAN, K.: Existence theorems and extreme solutions for inequalities concerning convex functions or linear transformations. Math. Z. **68**, 205—216 (1957).

The original paper containing the Neyman-Pearson lemma is

17. NEYMAN, J., and E. PEARSON: On the problem of most efficient tests of statistical hypotheses. Phil. Trans. Roy. Soc. (A) **231**, 289—337 (1933).

In connection with the method of orthogonal projection, see

18. WEYL, H.: The method of orthogonal projection in potential theory. Duke Math. J. **7**, 411—444 (1940).
19. LAX, P. D.: A remark on the method of orthogonal projection. Comm. Pure Appl. Math. **4**, 457—464 (1951).

A recent paper on the fundamental lemma is

20. GRAVES, L. M.: Extensions of the lemma of HAAR in the calculus of variations. Bull. Am. Math. Soc. **65**, 319—321 (1959).

See also

21. BERWALD, L.: Über HAARs Verallgemeinerung das Lemmas von DU BOIS-REYMOND und verwandte Sätze. Acta Math. **79**, 39—49 (1947).

The books

22. SCHWARTZ, L.: Théorie des distributions, I, II. Actual Sci. Ind. 1091, 1122. Paris: Hermann et Cie. 1950, 1951,

contain numerous references to the pioneer work of BOCHNER and FRIEDRICHS.

§ 2. A recent paper devoted to this topic is

1. URSELL, H. D.: Inequalities between sums of powers. Proc. London Math. Soc. **9**, 432—450 (1959).

§ 3. For the fundamentals of the theory of convexity, we refer the reader to [8.1] in Chapter 1 and to

1. EGGLESTON, H. G.: Convexity. Cambridge Tracts, no. 47. London: Cambridge University Press 1958.

We shall employ only some of the simplest ideas here. Many further methods are described and used in the papers by ROSENBLOOM cited above in [1.1].

§ 5.

1. BELLMAN, R., and D. BLACKWELL: On moment spaces. Ann. of Math. **54**, 272—274 (1951).
2. ROYDEN, H.: Bounds for distribution functions when the first n moment are given. Ann. Math. Stat. **24**, 361—376 (1953).

§ 7. Results of this nature were first given by F. RIESZ. See

1. RIESZ, F.: Untersuchungen über Systems integrierbarer Funktionen. Math. Ann. **69**, 449—497 (1910),
2. HELLY, E.: Über lineare Funktionaloperationen. Wiener Ber. **121**, 265—297 (1912),

and the book

3. RIESZ, F.: Les systemes d'equations linéaires a une infinité d'inconnues. Paris: Gauthier-Villars 1913.

For the usual proof, see BANACH [1.12]. For an application, see

4. BOAS, R. P.: A general moment problem. Am. J. Math. **63**, 361—370 (1941).

§ 8. It is very interesting to see how moment problems are so closely associated with oscillation properties of linear combinations of the basis functions, and thus intimately bound to problems of best approximation of the type initiated by ČEBYSHEV. For particular results, see

1. BERNSTEIN, S.: Leçons sur les propriétés extrémales et la meileure approximation des fonctions analytiques d'une variable réelle. Paris: Gauthier-Villars 1926.
2. WALSH, J. L.: Interpolation and approximation by rational functions in the complex domain. American Mathematical Society Colloquium Publications, vol. 20. New York: American Mathematical Society 1935.
3. DRESHER, M., and S. KARLIN: Solutions of convex games as fixed points. Ann. of Math. Studies, no. **28**, 75—86. Princeton, N. J.: Princeton University Press 1953.

A number of further results will be found in the book by KARLIN ([38.15] in Chapter 2).

§ 9.

1. WESTON, J. D.: A note on the extension of linear functionals. Am. Math. Monthly **67**, 444—445 (1960).

§ 10.

1. SZEGÖ, G.: Orthogonal polynomials. American Mathematical Society Colloquium Publications, vol. 23. New York: American Mathematical Society 1939.

§ 12.

1. FISCHER, E.: Über das Carathéodorysche Problem, Potenzreihen mit positivem reellen Teil betreffend. Rend. Circ. Mat. Palermo **32**, 240—256 (1911).

§ 13.

1. PICARD, E.: Mémoire sur les fonctions entières. Ann. Sci. École Norm. Supér. (2) **9**, 147—166 (1880).
2. BOREL, É.: Démonstration élémentaire d'un théorème de M. PICARD sur les fontions entrères. Compt. Rend. (Paris) **122**, 1045—1048 (1896).
3. LANDAU, E.: Über eine Verallgemeinerung des Picardschen Satzes. Sitz.ber. preuß. Akad. Wiss. **1904**, 1118—1133.
4. FEJÉR, L.: Über trigonometrische Polynome. J. Reine Angew. Math. **146**, 53—82 (1915).
5. CARATHÉODORY, C.: Über den Variabilitätsbereich der Fourierschen Konstanten von positiven harmonischen Funktionen. Rend. Circ. Mat. Palermo **32**, 193—217 (1911).
6. CARATHÉODORY, C., and L. FEJÉR: Über den Zusammenhang der Extremum von harmonischen Funktionen mit ihren Koeffizienten und über den Picard-Landauschen Satz. Rend. Circ. Mat. Palermo **32**, 218—239 (1911).
7. HERGLOTZ, G.: Über Potenzreihen mit positivem, reellem Teil im Einheitskreis. Sitz.ber. Sachs. Akad. Wiss. **63**, 501—511 (1911).

For a discussion of the Picard theorem and its generalizations from the function-theoretic point of view, see

8. BIEBERBACH, L.: Lehrbuch der Funktionentheorie, bd. 2, p. 217ff. Berlin: B. G. Teubner 1931.

Borel studied the Picard problem from the standpoint of solutions of the functional equation

$$e^{f(z)} + 1 = e^{g(z)},$$

where $f(z)$ and $g(z)$ are entire functions. Extensive generalizations exist along these lines.

9. LANDAU, G.: Darstellung und Begründung einiger neuerer Ergebnisse der Funktionentheorie. 2nd ed., chapter 7. Ergebnisse der Math. Berlin: J. Springer-Verlag 1929.

References to further papers by SCHUR, RIESZ, FROBENIUS, SZEGÖ, and TOEPLITZ will be found in RIESZ [7.1] and in

10. KAC, M., L. MURDOCK, and G. SZEGÖ: On the eigenvalues of certain hermitian forms. J. Rat. Mech. Analysis 9, 767—800 (1953).

These results, together with extensions and applications to probability theory, will be found in the book

11. GRENANDER, U., and G. SZEGÖ: Toeplitz forms and their applications. Berkeley and Los Angeles: University of California Press 1958.

A great deal of the interest in Toeplitz matrices stems from their similarity to certain matrices occurring in the study of various problems in statistical mechanics. In particular, there are close connections with the Ising problem in order-disorder theory. See

12. MARADUDIN, A., and G. H. WEISS: The disordered lattice problem, a review. J. Soc. Ind. Appl. Math. 6, 302—329 (1958).

§ 14. An excellent exposition of the theory of positive definite sequences and functions is to be found in the monograph [1.5] by KY FAN. See also

1. FAN, K.: On positive definite sequences. Ann. of Math. 47, 593—607 (1946).
2. LEVITAN, B.: A generalization of positive definiteness and a generalization of almost-periodic functions. Doklady Akad. Nauk SSSR 1947, 1593—1596.

§ 15. This theorem of Bochner occupies a crucial position in analysis. See the monograph [1.5] by KY FAN for a detailed discussion of this result and its antecedents.

1. LOOMIS, L.: An introduction to abstract harmonic analysis. New York: D. Van Nostrand Co. 1953.
2. WEIL, A.: L'integration dans les groupes topologiques et ses applications. Actual. Sci. Industr. 869. Paris: Hermann et Cie. 1940.

Some more recent results may be found in

3. DEVINATZ, A.: Some contributions to the extension of positive definite functions. Storrs, Conn.: University of Connecticut. OSR-TN-55-421, 1955.

For connections between the completely monotone functions of HAUSDORFF-BERNSTEIN-WIDDER and positive definite functions, see the books by FAN [1.5] and WIDDER [1.6].

4. BOCHNER, S.: Hilbert distances and positive functions, Ann. of Math. 42, 647—656 (1941).
5. SCHOENBERG, I. J.: Metric spaces and positive definite functions. Trans. Am. Math. Soc. 44, 522—536 (1938).
6. — and J. VON NEUMAN: Fourier integrals and metric geometry. Trans. Am. Math. Soc. 50, 226—251 (1941).
7. SZ.-NAGY, B., and A. KORÁNYI: Relation d'une problème de NEVANLINNA et PICK avec la théorie des operateurs de l'espace Hilbertian. Acta Math. Hung. 7, 295—305 (1956).
8. WEYL, H.: Über das Pick-Nevanlinnasche Interpolationsproblem und sein infinitesimales Analogen. Ann. of Math. 36, 230—254 (1935).

§ 16. The theory of reproducing kernels is a natural outgrowth of the integral equations of mathematical physics derived from the GREEN's function concept. Both theories contribute to each other. See

1. ARONSZAJN, N.: The theory of reproducing kernels. Trans. Am. Math. Soc. **68**, 337—404 (1950).
2. — and K. T. SMITH: Characterization of positive reproducing kernels. Applications to GREEN's functions. Am. J. Math. **79**, 611—622 (1957).

§ 17.

1. LIAPOUNOFF, A.: Sur les fonctions-vecteurs complètement additives. Bull. Acad. Sci. URSS Ser. Math. **4**, 465—478 (1940).

§ 18. The result was first given in

1. LANDAU, E.: Über einen Konvergenzsatz. Gott. Nachr. **1907**, 25—27,

and may also be found in the book by BANACH [1.1]. The result of ABEL may be found in the book by HARDY-Littlewood-Pólya ([1.1] of Chapter 1), Theorem 162, p. 120.

§ 19.

1. BELLMAN, R.: On an application of the Banach-Steinhaus theorem to the study of the boundedness of solutions of nonlinear differential and difference equations. Ann. of Math. **49**, 515—522 (1948).
2. MASSERA, J. L., and J. J. SCHÄFFER: Linear differential equations and functional analysis, I. Ann. of Math. **67**, 517—573 (1958).
3. — — Linear differential equations and functional analysis, II. Ann. of Math. **69**, 88—104 (1959).
4. CESARI, L.: Asymptotic behavior and stability problems in ordinary differential equations, p. 102. Ergebnisse der Math. Berlin: J. Springer-Verlag 1959.
5. CORDUNEANU, C.: Sur un théorème de Perron. Analele Stiintifice ale Univ. Jasi. **5**, 33—36 (1959).

§ 20.

1. GALE, D.: Theory of linear economic models. New York: McGraw-Hill Book Co., Inc. 1960.
2. WEYL, H.: The elementary theory of convex polyhedra. Ann. Math. Studies 24, 3—18. Princeton, N. J.: Princeton University Press 1950.

§ 21.

1. MINKOWSKI, H.: Zur Theorie der Einheiten in den algebraischen Zahlkörpern. Gesammelte Abhandlungen, I. Leipzig: B. G. Teubner 1911.
2. FARKAS, J.: Über die Theorie der einfachen Ungleichungen. J. Reine Angew. Math. **124**, 1—27 (1901).
3. STIEMKE, E.: Über positive Lösungen homogener linearer Gleichungen. Math. Ann. **76**, 340—342 (1915).
4. DINES, L. L.: Systems of linear inequalities. Ann. of Math. **20**, 191—199 (1919).
5. CARVER, W. B.: Systems of linear inequalities. Ann. of Math. **23**, 212—220 (1920—1922).
6. MOTZKIN, T.: Beiträge zur Theorie der linearen Ungleichungen. Diss., Basel; Azriel, Jerusalem 1936.
7. ČERNIKOV, S. N.: Systems of linear inequalities. Uspekhi Matem. Nauk, N. S. **8**, 7—73 (1953).

See also the book by MORGENSTERN ([42.2] in Chapter 2), particularly the chapters by WONG and WOODBURY. For the simplex method of DANTZIG, used to obtain numerical solutions, see

8. DANTZIG, G. B., A. ORDEN and P. WOLFE: The generalized simplex method for minimizing a linear form under linear inequality restraints. Pacific J. Math. **5**, 183—195 (1955).
9. FORD, L. R., J., and D. R. FULKERSON: A simple algorithm for finding maximal network flows and an application to the Hitchcock problem. Canad. J. Math. **9**, 210—218 (1957).

Further references will be found in

10. RILEY, V., and S. I. GASS: Linear programming and associated techniques. Baltimore, Md.: Operations Research Office, John Hopkins University 1958.
11. LEHMAN, S.: On confirmation and rational betting. J. Symbolic Logic **20**, 251—262 (1955).
12. NIKAIDÔ, H.: Sur un théorème de Stiemke. Osaka, Japan: Inst. Social Economic Research, Osaka University 1961.

§ 22.

1. GALE, D., H. KUHN and A. W. TUCKER: Nonlinear programming. Second Berkeley Symposium on Mathematical Statistics and Probability. Berkeley and Los Angeles: University of California Press 1951.
2. DUFFIN, R. J.: Infinite programs. Pittsburgh: Department of Mathematics, University of Pittsburgh, Technical Report No. 20, 1955.

The paper by LEHMAN contains a continuous version of the simplex method.

3. LEHMAN, S.: On the continuous simplex method. Santa Monica, Calif.: The RAND Corporation, Research Memorandum RM-1386, 1954.

The problems he treats arise in the discussion of continuous production processes; see [7.1] and [19.1] in Chapter 1, and

4. BELLMAN, R., and S. LEHMAN: Studies on bottleneck problems in a production processes. Santa Monica, Calif.: The RAND Corporation, Paper P-492, 1954.

§ 23. The classic work in the theory of games is

1. NEUMANN, J. VON, and O. MORGENSTERN: Theory of games and economic behavior. Princeton, N. J.: Princeton University Press 1953.

The study of two-person competitive processes was begun independently by BOREL and VON NEUMANN. See

2. BOREL, É.: Applications aux jeux de hasard. Traité du calcul des probabilities et de ses applications. Paris: Gauthier-Villars 1913,

and the translations (by L. J. SAVAGE) of BOREL's work, in Econometrica **21**, 97—117 (1953). The most readable and entertaining introductory account is

3. WILLIAMS, J. D.: The compleat strategyst. New York: McGraw-Hill Book Co., Inc. 1954.
4. LOOMIS, L. H.: On a theorem of VON NEUMANN. Proc. Nat. Acad. Sci. USA **32**, 213—215 (1946).
5. ROBINSON, J.: An iterative method for solving a game. Ann. of Math. **54**, 296—301 (1951).
6. GROSS, O.: A mechanical proof of the min-max theorem. Santa Monica, Calif.: The RAND Corporation, Research Memorandum RM-2320, 1960.
7. NIKAIDÔ, H.: Coincidence and some systems of inequalities. J. Math. Soc. Japan 11, 354—373 (1959).
8. NIKAIDÔ, H.: On a method of proof for the minimax theorem. Proc. Am. Math. Soc. 10, 205—212 (1959).
9. NEWMAN, J.: Another proof of the minimax theorem. Proc. Am. Math. Soc. **11**, 692—693 (1960).

§ 24. The proof of Theorem **15** is contained in an unpublished paper,

1. R. BELLMAN and M. SHIFFMAN, On the min-max of

$$\int_0^1 f(x)\, a(x)\, d(x)\, dt(x)\,,$$

and subsequently in Chapter 12 of the monograph [19.1] in Chapter 1.

§ 25. This amazingly powerful method is of rather late development. As mentioned in the text, a part of the fundamental idea was presented by ZAREMBA, whose work also anticipated that of BERGMAN-SCHIFFER and PICONE-FISCHERA concerning projection into the space of harmonic functions. The full result (obtained independently by CHEVALLEY and HERBRAND, but not published) was first given by WEYL [1.18]; for applications and further references, see

1. KACZMARZ, S., and H. STEINHAUS: Theorie der Orthogonalreihen. Monografje Matematyczne VI. Warsaw: Z Subwencji Funduszu Kultury Narodowej 1935.
2. FORSYTHE, G. E., and P. C. ROSENBLOOM: Numerical analysis and partial differential equations. Surveys in Applied Mathematics. New York: John Wiley and Sons 1958.

§ 26. See the results of Friedrichs referred to in the chapter on the calculus of variations in [26.1] of Chapter 2 and

1. LAX, P. D.: Reciprocal extremal problems in function theory. Comm. Pure Appl. Math. 8, 437—453 (1955).
2. ROGOSINSKI, W. W., and H. S. SHAPIRO: On certain extremum problems for analytic functions. Acta Math. 90, 287—318 (1953).

Chapter 4

On the Positivity of Operators

§ 1. Introduction

In this chapter, we shall explore the following theme: "Given a set of functions $\{u\}$ satisfying certain side conditions, and an operator L that can be applied to the functions of this set, determine when the inequality

$$L(u) \geqq 0 \tag{1}$$

implies that $u \geqq 0$." An operator will be said to be "positive" if this condition is satisfied, although it might be more reasonable to call the inverse operator L^{-1} positive. We shall focus our attention upon ordinary differential and partial differential operators.

Problems of this general nature were studied extensively by ČAPLYGIN [1], who thereby developed techniques for obtaining useful approximations to the solutions of differential equations of complicated nature. We shall discuss some of his results below, together with some later refinements.

We shall begin our discussion with first-order ordinary differential operators and then turn to closely related results of GRONWALL [2], BELLMAN [3], BIHARI [4], and LANGENHOP [5], of importance in the study of the existence, uniqueness, and stability of solutions of ordinary differential equations.

The problem for higher-order linear differential operators is intimately connected with characteristic values and two-point boundary-value questions. To begin with, we shall consider the second-order ordinary differential inequality

$$u'' + g(x)\, u \geqq 0, \quad u(0) = u(1) = 0\,, \tag{2}$$

using a variety of techniques; some of these carry over to partial differential operators, while others do not.

We shall first present some very simple arguments of the type that are used in the study of the oscillation and nonoscillation of solutions of second-order differential equations. Extension of these arguments can be used to establish a number of "maximum principles" for solutions of partial differential equations. Following this, we shall give a proof of Čaplygin [1], based on an integral identity, and a proof of Bellman [12.1] that uses a closely related variational argument. We shall refer to other proofs by Petrov [15.1] and Wilkins [15.2], based on the Riccati equation, and finally to a proof employing factorization of the operator, as suggested by work of Poincaré [6].

This method of factorization of the operator was used by Pólya [7] to study generalized Rolle's theorems and interpolation problems for n-th order linear differential operators. The results of Pólya were extended to partial differential operators by a number of mathematicians. Full references and a unified presentation will be found in a paper by Hartman and Wintner [8].

The study of these questions brings us in a very natural way to the concept of generalized convexity and the investigations of Beckenbach [17.1], Bonsall [17.4], Peixoto [17.5—17.7], Reid [17.11], Hartman [17.12], and others. These results make precise some of the initial results of Čaplygin mentioned above.

Let us also note that these questions are intimately connected with generalizations of the Taylor expansion, along the lines of Pólya [7], Delsarte [20.1], Petersson [16.1], Levitan [20.2], and Widder [20.3].

We also present some positivity results for vector-matrix differential equations of the form

$$\frac{dx}{dt} = A(t)\,x, \quad x(0) = c\,, \tag{3}$$

where $x(t)$ is a vector function and $A(t)$ a matrix function.

Next we turn our attention to partial differential operators, considering only the classical equations

$$\begin{aligned} u_t - u_{xx} - u_{yy} - u_{zz} - q(x, y, z)\,u &= 0\,, \\ u_{xx} + u_{yy} + u_{zz} - q(x, y, z)\,u &= 0\,, \end{aligned} \tag{4}$$

although the more general Beltrami equations can be treated by several of the techniques presented. The positivity of the associated operators is, of course, equivalent to the positivity of the Green's function. This topic has been approached in a number of different ways; see Aronszajn and Smith ([16.2] of Chapter 3) for references. As we shall see, a number of the methods applicable to the study of ordinary differential operators carry over.

A very elegant and penetrating way to establish the positivity of GREEN's functions is to exhibit the probabilistic interpretation of these functions. An excellent exposition of this modern approach is given by KAC [9]. For an analogous use of probability theory to derive the positivity of certain expressions, see KARLIN and MCGREGOR ([37.9] of Chapter 2).

Our excursion into this field of partial differential inequalities will be brief. We shall only mention the fundamental positivity results of HAAR [29.1], WESTPHAL [29.2], PRODI [29.3], and MLAK [29.5], of importance in the stability theory of parabolic partial differential equations. These are, of course, only particular results in the domain of "maximum principles" of partial differential equation theory; see NURENBERG [21.1], WEINBERGER [21.2], and PUCCI [29.4], where many other references will be found.

We have felt that these, as well as the extensions of the classical Sturmian theory for ordinary differential equations, due to HARTMAN and WINTNER [8], and to REDHEFFER [35.9], belong more to the theory of partial differential equations than to the theory of inequalities, and so have omitted any discussion of them.

Finally, we have indicated briefly how positive operators may be used in the study of nonlinear functional equations, in connection with the quasi-linearization techniques mentioned in Chapter 1 and 2.

It is clear from the foregoing how difficult it would be to trace all the tributaries of this mainstream of analysis, the theory of positive operators. We hope that we have sufficiently indicated some of the principal currents, and given sufficient references so that the interested reader can journey on his own. Let us observe in passing that the results we have presented can all be greatly extended by replacing the concept of a positive operator by that of a variation-diminishing operator, along the lines of the work of SCHOENBERG [10] and others to which we have previously referred in [48.1], [48.2], and [48.10] of Chapter 2.

§ 2. First-order Linear Differential Equations

Our first result concerning differential operators is the following:

Theorem 1. *If the linear differential equation*

$$\frac{du}{dt} = a(t)\, u, \quad u(0) = c\,, \tag{1}$$

and the linear differential inequality

$$\frac{dv}{dt} \geqq a(t)\, v, \quad v(0) = c\,, \tag{2}$$

are both valid for $0 \leqq t \leqq T$, *then*

$$v(t) \geqq u(t), \quad 0 \leqq t \leqq T. \tag{3}$$

Proof. The proof is an immediate consequence of the fact that the solution of the linear inhomogeneous equation

$$\frac{dv}{dt} = a(t)\, v + f(t), \quad v(0) = c, \tag{4}$$

has the form

$$v = c e^{\int_0^t a(s)\, ds} + \int_0^t e^{\int_r^t a(s)\, ds} f(r)\, dr. \tag{5}$$

The positivity of the kernel

$$e^{\int_r^t a(s)\, ds}$$

is the key to the result.

§ 3. Discussion

The result presented in the preceding section, although quite simple to prove, is important for two reasons. In the first place, it illustrates the type of result we wish to establish; and in the second place, it sets the following pattern for a type of proof that can be used in many situations.

The inequality $L(u) \geqq 0$, where L is a linear differential operator, is converted into the inhomogeneous equation

$$L(u) = f(p), \tag{1}$$

where $f(p)$ is nonnegative. Solving for u, we obtain a relation of the type

$$u = T(f). \tag{2}$$

The problem has then been converted to that of studying the positivity properties of the operator T, or, equivalently, those of the Green's function associated with L. Occasionally, as in § 2, these properties are apparent, but in the majority of cases an artifice of one type or another is required to complete the proof.

§ 4. A Fundamental Result in Stability Theory

Closely related to the foregoing Theorem 1 is the following result of Bellman [1.3].

Theorem 2. *If the functions* $g(t)$ *and* $u(t)$ *are nonnegative for* $t \geqq 0$, *and if* $c \geqq 0$, *then the inequality*

$$u(t) \leqq c + \int_0^t g(s)\, u(s)\, ds, \quad t \geqq 0, \tag{1}$$

implies that

$$u(t) \leqq c e^{\int_0^t g(s)\,ds}, \quad t \geqq 0. \tag{2}$$

This result may be established either directly or by means of the technique of § 2; see also GRONWALL [1.2] and GUILIANO [3]. In view of the frequent occurrence of the result, let us give a quick proof [1]:

From (1), we have

$$\frac{u(t)\, g(t)}{c + \int_0^t g(s)\, u(s)\, ds} \leqq g(t), \tag{3}$$

whence, integrating from 0 to t, we obtain

$$\log \left[c + \int_0^t g(s)\, u(s)\, ds\right] - \log c \leqq \int_0^t g(s)\, ds. \tag{4}$$

This yields

$$c + \int_0^t g(s)\, u(s)\, ds \leqq c e^{\int_0^t g(s)\,ds}. \tag{5}$$

The desired inequality (2) follows from (1) and (5).

Various applications of this result to the study of stability of the solutions of linear and nonlinear differential equations may be found in BELLMAN [1]. Numerous applications to existence and uniqueness theory of differential equations may be found in NEMYCKII-STEPANOV [2], BIHARI [1.4], and LANGENHOP [1.5]; see also LAX [4] for an application to the establishment of *a priori* bounds for the solutions of a class of partial differential equations.

§ 5. Inequalities of Bihari-Langenhop

The following generalization of the foregoing Theorem 2 was obtained by BIHARI [1.4].

Theorem 3. *If $k, m \geqq 0$, and $g(s)$ is positive for $s > 0$, then the inequality*

$$u(t) \leqq k + m \int_a^t v(s)\, g(u(s))\, ds, \quad a \leqq t \leqq b, \tag{1}$$

implies that

$$u(t) \leqq G^{-1}\left(G(k) + m \int_a^t v(s)\, ds\right), \tag{2}$$

where

$$G(u) = \int_{u_0}^{u} \frac{dt}{g(t)}, \quad u > u_0 > 0. \tag{3}$$

This result was used by BIHARI in the way mentioned above. Closely related is the following result due to LANGENHOP [1.5]. Let

(a) x be a real variable and z and F be finite-dimensional complex vectors with n components z_i and F_i, respectively; (4)

(b) F be continuous in (x, z) for all z and all $x \in [a, b]$, i. e., $a \leqq x \leqq b$ with $a < b$;

(c) for some norm, say $||z|| = \sum_{i=1}^{n} |z_i|$, F satisfy

$$||F(x, z)|| \leqq v(x)\, g(||z||)\,,$$

where

$v(x)$ is continuous, $v(x) \geqq 0$ for $x \in [a, b]$,
$g(u)$ is continuous and nondecreasing for $u \geqq 0$,
and $g(u) > 0$ for $u > 0$.

If $z(x)$ is continuous, and is a solution of $dz/dx = F(x, z)$ for $x \in [a, b]$, where F satisfies the conditions above, then, for $x \in [a, b]$, $z(x)$ satisfies the inequality

$$|z(x)| \geqq G^{-1}\left(G(|z(a)|) - \int_a^x v(s)\, ds\right),$$

where

$$G(u) = \int_{u_0}^{u} [g(t)]^{-1}\, dt\,, \quad u_0 \geqq 0\,,$$

for all $x \in [a, b]$ for which $G(|z(a)|) = \int_a^x v(s)\, ds$ is in the domain of G^{-1}.

Whereas the Bihari result furnishes upper bounds, the Langenhop result yields lower bounds.

§ 6. Matrix Analogues

Let us now discuss a vector-matrix analogue of Theorem 1. Consider the vector-matrix inequality

$$\frac{dx}{dt} \geqq A(t)\, x, \quad x(0) = c\,, \tag{1}$$

where $A(t)$ is a matrix function, and x a vector function, of order n. In the one-dimensional case, this inequality implies that x is bounded from below by the solution of the equation

$$\frac{dy}{dt} = A(t)\, y\,, \quad y(0) = c\,. \tag{2}$$

In the multidimensional case, no such uniform result holds. For the case in which A is a constant, however, there is a simple result of interest.

Theorem 4. *Let $A = (a_{ij})$ be a constant matrix. A necessary and sufficient condition that a solution of the inequality*

$$\frac{dx}{dt} \geqq Ax, \quad x(0) = c, \tag{3}$$

be bounded from below over $t \geqq 0$ by the solution of the equation

$$\frac{dy}{dt} = Ay, \quad y(0) = c, \tag{4}$$

is that

$$a_{ij} \geqq 0, \quad i \neq j, \quad a_{ii} \text{ real.} \tag{5}$$

Proof. Since the solution of the inhomogeneous equation

$$\frac{dx}{dt} = Ax + f(t), \quad x(0) = c, \tag{6}$$

has the form

$$x = e^{At} c + \int_0^t e^{A(t-s)} f(s)\, ds, \tag{7}$$

where e^{At} is the matrix exponential (see [43.1] in Chapter 2), we see that it is essential to determine when the elements of e^{At} are nonnegative for $t \geqq 0$.

The expansion

$$e^{At} = I + At + \cdots, \tag{8}$$

for small positive t, shows that the condition in (5) is necessary. To see that it is sufficient, we can use either the system of differential equations

$$\frac{dy_i}{dt} = \sum_{j=1}^{n} a_{ij} y_j, \quad i = 1, 2, \ldots, n, \tag{9}$$

$$y_i(0) = c_i,$$

as indicated in § 8, below, or the identity

$$e^{At} = (e^{At/N})^N. \tag{10}$$

Let us use the identity first. We can assume that $a_{ij} > 0$, $i \neq j$, since the result for $a_{ij} \geqq 0$ can then be obtained via a limiting procedure. For fixed t, we see from the expression (8) that $e^{At/N}$ is positive, in the sense that all elements are positive, for N sufficiently large. Since the product of positive matrices is positive, it follows that e^{At} is positive if $a_{ij} > 0$, $i \neq j$. This proof is due to KARLIN.

Theorem 4 is intimately connected with the Perron theorem concerning positive matrices, and with its extension, given in Chapter 2; see also [38.3] in Chapter 2, where the Perron theorem is derived from results pertaining to linear differential equations. Let us note that the result is intuitively clear once the probabilistic or economic origin of equation (9) is made clear; cf. ROMANOVSKY [1] and OPIAL [2].

§ 7. A Proof by Taussky

The following short unpublished proof by TAUSSKY possesses the merit of being applicable to the derivation of a number of results for input-output matrices directly from the corresponding results for positive matrices.

Let k be a scalar. Then

$$e^{At} = e^{(A+kI)t} e^{-kIt}. \tag{1}$$

If k is chosen large enough so that $A + kI$ is nonnegative, we see that $e^{(A+kI)t}$ will be a nonnegative matrix. Since $e^{-kIt} \geqq 0$, it follows that e^{At} is nonnegative.

§ 8. Variable Matrix

If $A(t)$ is a variable matrix, the solution of the linear inhomogeneous equation

$$\frac{dx}{dt} = A(t)\, x + f(t), \quad x(0) = c\,, \tag{1}$$

may be written

$$x = Y(t)\, c + \int_0^t Y(t)\, Y^{-1}(s)\, f(s)\, ds\,, \tag{2}$$

where $Y(t)$ is the solution of the matrix equation

$$\frac{dX}{dt} = A(t)\, X, \quad X(0) = I\,. \tag{3}$$

From this representation, we see that a necessary and sufficient condition that

$$\frac{dx}{dt} = A(t)\, x, \quad x(0) \geqq 0\,, \tag{4}$$

imply $x \geqq 0$ for $t \geqq 0$ is that $Y(t) \geqq 0$, while a necessary and sufficient condition that

$$\frac{dx}{dt} \geqq A(t)\, x, \quad x(0) = 0\,,$$

imply $x \geqq 0$ for $t \geqq 0$ is that $Y(t)\, Y^{-1}(s) \geqq 0$ for $t \geqq s \geqq 0$.

These two conditions do not seem to be equivalent as they are in the case in which $A(t)$ is constant. A simple sufficient condition that $Y(t) \geqq 0$ for $t \geqq 0$ is $a_{ij}(t) \geqq 0$, $i \neq j$, as may be seen from the following simple argument. Suppose, without loss of generality, that all of the $x_i(0)$ are positive, and let $x_1(t)$ be the component that is zero for the first time at a point t_0. Then at t_0, we have

$$\frac{dx_1}{dt} = a_{12} x_2 + \cdots + a_{1n} x_n > 0\,, \tag{5}$$

a contradiction to the fact that x_1 must be decreasing as t approaches t_0.

This sufficiency condition is important in the study of the equation

$$\frac{dx}{dt} = \max_q \left[A(q)\, x + b(q)\right], \quad x(0) = c\,, \tag{6}$$

arising in the theory of MARKOVIAN decision processes ([6.1], [7.1] of Chapter 1, and [44.2] of Chapter 2), and in the use of quasi-linearization techniques.

§ 9. Discussion

The results in § 8 concern only one particular problem arising in the study of Equation (8.1). One may also ask for necessary and sufficient conditions that the solution $x(t)$ be bounded for $t \geqq 0$ whenever $f(t)$ is bounded for $t \geqq 0$, or that $x(t) \in L^2(0, \infty)$ whenever $f(t) \in L^2(0, \infty)$, and so on. Generally, one can ask when $f \in S$ implies $x \in S'$, where S and S' are Banach spaces. This study was initiated by PERRON [1], who used direct methods, and continued through Banach-space techniques by BELLMAN ([19.1] of Chapter 3) and MASSERA [2], [3]. The work of Massera covers operator equations as well.

§ 10. A Result of Čaplygin

Let us begin our study of the second-order linear differential operator by establishing a slight extension of a result due to ČAPLYGIN [1.1]. The first method we employ is widely used in the study of the oscillation of the solutions of linear and nonlinear second-order differential equations. In many cases, it can be carried over *in toto* to establish analogous results for partial differential equations.

Theorem 5. *If*

$$\begin{aligned}
&\text{(a)} \quad u'' + p(t)\, u' - q(t)\, u > 0, \quad t \geqq 0\,,\\
&\text{(b)} \quad v'' + p(t)\, v' - q(t)\, v = 0, \quad t \geqq 0\,,\\
&\text{(c)} \quad q(t) \geqq 0, \quad t \geqq 0\,,\\
&\text{(d)} \quad u(0) = v(0), \quad u'(0) = v'(0)\,,
\end{aligned} \tag{1}$$

then $u > v$ *for* $t \geqq 0$.

Proof. Subtracting, we have

$$w'' + p(t)\, w' - q(t)\, w > 0\,, \tag{2}$$

where $w = u - v$, with $w(0) = w'(0) = 0$. It follows that $w > 0$ in some initial interval $(0, t_0]$. Suppose that w eventually becomes negative, so that w must have a local maximum at some point t_1. At this point, we would have $w' = 0$, $w > 0$, and therefore, by (2), $w'' > 0$. This, however, contradicts the assumption that t_1 is a local maximum.

ČAPLYGIN's proof, for the special case in which $p(t) = 0$, depends on an important identity,

$$\int_0^a u\,(u'' - qu)\,dt = [uu']_0^a - \int_0^a (u'^2 + qu^2)\,dt\,. \tag{3}$$

Let a be the first positive value of t for which $u = 0$. Then, if $u(0) = 0$, we have

$$\int_0^a u\,(u'' - qu)\,dt = -\int_0^a (u'^2 + qu^2)\,dt\,. \tag{4}$$

If $u'' - qu > 0$ in [0, a], and $u \geqq 0$ in this same interval, we clearly have a contradiction. Hence, such a point a does not exist.

For a study of the converse problem of dertermining when positive operators have the foregoing form, see FELLER [1].

§ 11. Finite Intervals

Analyzing the foregoing proof, we see that the condition $q \geqq 0$ can be considerably relaxed, provided that we confine our attention to finite intervals. This is a consequence of the fact that the inequality

$$\int_0^a (u'^2 + qu^2)\,dt \geqq 0 \tag{1}$$

can hold for negative q, provided that $|q|$ is not too large. This is a corollary of the Wirtinger inequality, which we shall discuss in detail in the following chapter, and which is itself a corollary of general Sturm-Liouville theory.

From (10.4) we readily obtain the following result:

Theorem 6. *If*

$$\begin{aligned} &\text{(a)}\quad u'' - qu \geqq 0, \quad 0 \leqq t \leqq a\,,\\ &\text{(b)}\quad u(0) = u(a) = 0\,,\\ &\text{(c)}\quad \int_0^a (u'^2 + qu^2)\,dt \geqq 0\,, \end{aligned} \tag{2}$$

then $u \leqq 0$ for $0 \leqq t \leqq a$.

As we shall see in § 12 of Chapter 5, a sufficient condition for (2c) to hold is

$$q(t) \geqq -\frac{\pi^2}{a^2} + \delta, \quad \delta > 0, \quad 0 \leqq t \leqq a. \tag{3}$$

The magic quantity π^2/a^2 is, of course, the first characteristic value associated with the Sturm-Liouville equation

$$\begin{aligned} &u'' + \lambda u = 0\,,\\ &u(0) = u(a) = 0\,. \end{aligned} \tag{4}$$

§ 12. Variational Proof

Let us now present a variational argument that can be used to obtain a number of further results concerning the GREEN's function associated with

$$\begin{aligned} &u'' + q(t)\,u = f(t)\,, \\ &u(0) = u(a) = 0\,. \end{aligned} \tag{1}$$

We change the sign of $q(t)$ in order to indicate that it is positive in many significant cases.

If $q(t) \equiv 0$, it is clear that the inequality $f(t) \geqq 0$ for $0 \leqq t \leqq a$ implies that $u(t) \leqq 0$ in this interval. A more precise result is the following.

Theorem 7. *If*

$$\begin{aligned} &\text{(a)} \quad q(t) \leqq \frac{\pi^2}{a^2} - d, \quad d > 0, \quad 0 \leqq t \leqq a\,, \\ &\text{(b)} \quad f(t) \geqq 0, \quad 0 \leqq t \leqq a\,, \end{aligned} \tag{2}$$

then $u(t) \leqq 0$, $0 \leqq t \leqq a$.

Proof. Consider the problem of minimizing the functional

$$J(u) = \int_0^a [u'^2 - q(t)\,u^2 + 2f(t)\,u]\,dt \tag{3}$$

over all u satisfying the constraints $u(0) = u(a) = 0$, and for which the integral exists.

Since

$$\int_0^a q(t)\,u^2\,dt \leqq \left(\frac{\pi^2}{a^2} - d\right) \int_0^a u^2\,dt \leqq \frac{(\pi^2/a^2 - d)}{\pi^2/a^2} \int_0^a u'^2\,dt\,, \tag{4}$$

by virtue of WIRTINGER's inequality mentioned above, it follows via standard variational arguments that the minimum of $J(u)$ exists and is furnished by a unique function u. The Euler variational equation for this function is precisely (1).

Thus, to demonstrate that the solution of equation (1) is nonpositive for $0 \leqq t \leqq 1$ whenever $f(t) \geqq 0$ in $[0,1]$, it is sufficient to show that the function that minimizes $J(u)$ is nonpositive whenever $f(t)$ is nonnegative.

This, however, is easily established. Assume that $[a_1, b_1]$ is an interval within $[0, a]$ with the property that $u(t) > 0$ in $[a_1, b_1]$. Replace $u(t)$ by a new function equal to $-u(t)$ in $[a_1, b_1]$, and preserving the old values elsewhere. Although this introduces a possible discontinuity into $u'(t)$, it does not affect L^2-integrability. This change does not affect the quadratic terms, and it diminishes the term involving $f(t)$. Consequently, if $f(t)$ is positive on a set of positive measure in $[a_1, b_1]$, we obtain a contradiction. See BELLMAN [1] and STIELTJES [2].

To establish the theorem, it is clearly sufficient to consider only functions $f(t)$ that are positive within $[0, a]$.

§ 13. Discussion

It is easy to see that the foregoing Theorem 7 is the best possible in the sense that if (12.2a) is violated, the conclusion is not necessarily true. As an instance of this, consider the solution of

$$\begin{aligned} u'' + ku &= \sin\frac{\pi t}{a}, \\ u(0) &= u(a) = 0, \end{aligned} \tag{1}$$

given by

$$u = \frac{\sin \pi t/a}{k - \pi^2/a^2}. \tag{2}$$

If $k > \pi^2/a^2$, then u is positive for $0 < t < a$, although $\sin \pi t/a \geqq 0$ in $[0, a]$.

As mentioned above, the quantity π^2/a^2 enters as the smallest characteristic value of the Sturm-Liouville equation

$$\begin{aligned} u'' + \lambda u &= 0, \quad 0 \leqq t \leqq a, \\ u(0) &= u(a) = 0. \end{aligned} \tag{3}$$

The positivity result of Theorem 7 will be obtained from another direction, through a quite different technique, in the following § 14.

Since the solution of

$$\begin{aligned} u'' + q(t)\, u &= f(t), \\ u(0) &= u(a) = 0, \end{aligned} \tag{4}$$

may be written in the form

$$u = \int_0^a k(t, s) f(s)\, ds, \tag{5}$$

where $k(t, s)$ is the GREEN's function associated with the equation *cum* boundary conditions, it is clear that any assertion concerning the solution of (4) for *all* nonnegative f is an assertion concerning the nonnegativity of $k(t, s)$.

Finally, let us note that the argument given above is similar to one used by STIELTJES [12.2] in the discussion of systems of linear equations and inverses of matrices. His result has been greatly extended; see [2.1] of Chapter 2 for many further references.

§ 14. Linear Differential Equations of Arbitrary Order

Let us now consider a more general version of the problem. Let L be a linear differential operator of order n,

$$L(u) = \frac{d^n u}{dt^n} + a_1(t)\frac{d^{n-1} u}{dt^{n-1}} + \cdots + a_n(t)\, u, \tag{1}$$

where the $a_i(t)$ are, let us say, continuous over an interval $[0, t_0]$, and let $u_1, u_2, \ldots, u_n$ represent n linearly independent solutions of $L(u) = 0$. Introduce the Wronskian determinants,

$$W_1(t) = u_1(t), \quad W_2(t) = \begin{vmatrix} u_1(t) & u_2(t) \\ u_1'(t) & u_2'(t) \end{vmatrix}, \ldots,$$
$$W_n(t) = \begin{vmatrix} u_1(t) & u_2(t) & \ldots u_n(t) \\ u_1'(t) & u_2'(t) & \ldots u_n'(t) \\ \vdots & & \\ u_1^{(n-1)}(t) & u_2^{(n-1)}(t) & \ldots u_n^{(n-1)}(t) \end{vmatrix}. \tag{2}$$

The WRONSKIAN of order n has the well-known evaluation

$$W_n(t) = e^{-\int_0^t a_1(s)\,ds}, \tag{3}$$

a result due to JACOBI, and we take $W_0(t) \equiv 1$.

Our results concerning the positivity of the operator L hinges upon the following interesting representation.

Theorem 8. *If the $W_i(t)$, $i = 1, 2, \ldots, n-1$, are positive in $[0, t_0]$, then, in this interval, we may write*

$$L = \frac{W_n}{W_{n-1}} \frac{d}{dt} \left[\frac{W_{n-1}^2}{W_{n-2} W_n} \cdots \frac{d}{dt} \left(\frac{W_2^2}{W_1 W_3} \frac{d}{dt} \left(\frac{W_1^2}{W_2} \right) \cdots \right) \right]. \tag{4}$$

Proof of this result may be found in PÓLYA and SZEGÖ ([44.1] of Chapter 1).

§ 15. A Positivity Result for Higher-order Linear Differential Operators

It is not difficult to obtain the following theorem from the foregoing result.

Theorem 9. *A sufficient condition that the inequality*

$$L(u) \geqq 0, \quad u(0) = u'(0) = \cdots = u^{(n-1)}(0) = 0, \tag{1}$$

for $0 \leqq t \leqq t_1$ imply that $u \geqq 0$ in $[0, t_1]$ is that there exist a set of linearly independent solutions $u_1, u_2, \ldots, u_n$ for which the $W_i(t)$ are positive in $[0, t_1]$.

This result generalizes corresponding results obtained for the case of second-order linear differential operators by PETROV [1] and WILKINS [2]. In the paper by WILKINS, the condition is stated in terms of the associated Riccati equation. Whether or not the condition is necessary in general, as it is in the case $n = 2$, is not immediately clear.

An extensive discussion of results of the foregoing nature is contained in the monograph by ČAPLYGIN [1.1], where analytic and geometric arguments are given, together with a number of applications

§ 16. Some Results of Pólya

The property used in Theorem 8, namely the assumption that there exist n linearly independent solutions $u_1, u_2, \ldots, u_n$ for which the WRONSKIANS $W_i(t)$ are positive in $[0, t_1]$, was called by PÓLYA [1.7] *property W*. Using this condition, he established the following results.

Theorem 10. *If the linear differential operator*

$$L(u) = u^{(n)} + a_1(t)\, u^{(n-1)} + \cdots + a_n(t) u \tag{1}$$

possesses property W, then given any function $f(t)$ defined over $[0, t_1]$, n times differentiable there, and vanishing at $n+1$ points in $[0, t_1]$, there exists an internal point s such that

$$L\,(u(s)) = 0\,. \tag{2}$$

This is an extensive generalization of ROLLE's Theorem.

Theorem 11. *If property W holds, than there exists one and only one solution of $L(u) = 0$ that takes on n given values at n given points of $[0, t_1]$.*

This is an existence and uniqueness proof for the Lagrange interpolatin theorem. See PETERSSON [1].

Theorem 12. *Assuming property W, let v be a solution of $L(v) = 0$, possessing the same values as a function u at n given points of $[0, t_1]$. Let $w(t)$ be the solution of $L(u) = 1$ that vanishes at these same n points. Then there is a point $s = s(t)$ in $[0, t_1]$, corresponding to each t in $[0, 1]$, such that*

$$u(t) = v(t) + w(t)\, L\left(u(s(t))\right). \tag{3}$$

This is a generalized mean-value theorem. The special case corresponding to the operator

$$u'' + u \tag{4}$$

was established by POINCARÉ [1.6] and furnished the stimulus for PÓLYA's investigations. For some further results, see HARTMAN [17.12] and the papers cited in §§ 19, 20 pertaining to generalized Taylor expansions.

An interesting corollary of these results is the fact that the determinant $|e^{x_i y_j}|$ satisfies

$$|e^{x_i y_j}| \neq 0 \tag{5}$$

if $x_1 < x_2 < \cdots < x_n$, $y_1 < y_2 < \cdots < y_n$. This type of result is also a consequence of the theory of variation-diminishing transformations, and, as pointed out by SCHOENBERG [1.9], there are many points of contact between the two studies.

§ 17. Generalized Convexity

Let us now, following the ideas of BECKENBACH [1], BECKENBACH and BING [2], VALIRON [3], BONSALL [4], PEIXOTO [5], [6], [7], MOTZKIN [8], TORNHEIM [9], CURTIS [10], REID [11], and HARTMAN [12], discuss briefly the concept of generalized convexity, as introduced by BECKENBACH [1], and its relevance to the foregoing topics.

A function $u(t)$ that is convex for $a \leqq t \leqq b$ may be described in the following fashion. Let $a < t_1 < t_2 < b$, and let $v(t)$ be the solution of the linear equation

$$\frac{d^2 u}{d t^2} = 0 \tag{1}$$

passing through the two points $(t_1, u(t_1))$, $(t_2, u(t_2))$. Then

$$u(t) \leqq v(t)\,, \tag{2}$$

for $t_1 \leqq t \leqq t_2$. See Fig. 6.

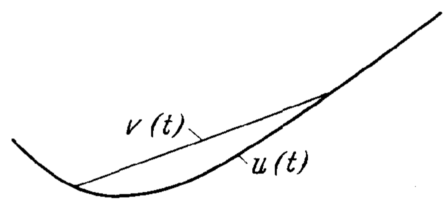

Fig. 6

To generalize the concept of convexity, we need merely generalize the differential operator under consideration. BONSALL [4] considered the second-order linear differential equation

$$u'' + p(t)\, u' + q(t)\, u = 0\,, \tag{3}$$

while PEIXOTO [7] used the more general nonlinear differential equation

$$u'' = g(t, u, u'). \tag{4}$$

The following result was established by PEIXOTO.

Theorem 13. *Assume that the equation in* (4) *possesses the following properties:*

(a) *$g(t, u, u')$ is continuous in t, u, and u' for $a < t < b, -\infty < u < \infty, -\infty < u' < \infty$.* (5)

(b) *Corresponding to any point $a < t_0 < b$, $-\infty < u_0 < \infty, -\infty < v_0 < \infty$, there is a unique solution of* (4) *satisfying the conditions $u(t_0) = u_0$, $u'(t_0) = v_0$, for $a < t < b$.*

(c) *Given any two points $a < t_i < b$, $-\infty < u_i < \infty$, $-\infty < v_i < \infty$, $i = 1, 2$, there is a unique solution of* (4) *satisfying the conditions $u(t_i) = u_i$, $u'(t_i) = v_i$, $i = 1, 2$.*

If $w(t)$ is a function with continuous second derivatives for $a < t < b$, then a necessary and sufficient condition that $w(t) \leqq u(t)$, $t_1 \leqq t \leqq t_2$, where u

is determined by the conditions of (5c) *and* $w(t_i) = u(t_i)$, $i = 1, 2$, *is that*

$$w'' \geqq g(t, w, w'), \quad a < t < b. \tag{6}$$

BONSALL [4] established the corresponding result for the second-order linear equation (3) without the requirement that w possess a continuous second derivative. In a subsequent paper [19], he discussed corresponding problems associated with partial differential operators. In this way, we enter the domain of subharmonic and superharmonic functions and their generalizations, a subject we shall discuss in our second volume on inequalities. See RIESZ [13], RADÓ [14], BECKENBACH and RADÓ [15], [16], BECKENBACH [17], READE [18], TAUTZ [20], BECKENBACH and JACKSON [21], JACKSON [22], [24], and INONE [23].

§ 18. Discussion

As indicated in the introduction to this chapter and also in § 15, ČAPLYGIN [1.1] studied the relation between functions satisfying the inequality

$$u^{(n)} - g(t, u, u', \ldots, u^{(n-1)}) > 0, \tag{1}$$

and functions satisfying the equation

$$v^{(n)} - g(t, v, v', \ldots, v^{(n-1)}) = 0. \tag{2}$$

Let $u^{(k)}(0) = v^{(k)}(0) = c_k$, for $k = 0, 1, \ldots, n-1$. Then, as pointed out by ČAPLYGIN, it is clear that there is some t-interval, $0 \leqq t \leqq t_0$, in which $u \geqq v$, for we may write

$$\begin{aligned} u &= c_0 + c_1 t + \cdots + \frac{c_{n-1} t^{n-1}}{(n-1)!} + \frac{u^{(n)}(0)\, t^n}{n!} + R_n, \\ v &= c_0 + c_1 t + \cdots + \frac{c_{n-1} t^{n-1}}{(n-1)!} + \frac{v^{(n)}(0)\, t^n}{n!} + S_n. \end{aligned} \tag{3}$$

Since $u^{(n)}(0) > v^{(n)}(0)$, by virtue of (1) and (2), and $|R_n| = O(t^{n+1})$, $|S_n| = O(t^{n+1})$ for small t, we see that our assertion is valid.

A number of applications of inequalities to the problem of obtaining upper and lower bounds for solutions of ordinary differential equations were given by ČAPLYGIN [1.1] and also by a number of other authors; see PÓLYA [1.7], BONSALL [17.4], PETROV [15.1], WILKINS [15.2], HARTMAN [17.12], and KLIMKO [1], ARTEMOV [2], and PARODI [3].

§ 19. The Generalized Mean-value Theorem of Hartman and Wintner

The mean-value theorem for linear differential operators presented in § 18 possesses various analogues for partial differential operators, as was indicated by PÓLYA [1]; see also BLEULER [2].

A generalized mean-value theorem, due to HARTMAN and WINTNER [1.8], which abstracts the essence of PÓLYA's results and methods, is the following.

Theorem 14. *Let $\{u(p)\}$ be a set of functions defined for p in a region R, and admitting a linear operator L. Let B be the boundary of R, and denote by S the set of functions satisfying a fixed boundary condition. Let the linear operator L possess the following two properties:*

$$\begin{aligned} &\text{(a)} \quad \text{There exists a solution of } L(v) = 1,\ p \in R, \text{ with } v \in S. \\ &\text{(b)} \quad \text{If } u \in S, \text{ and } L(u) \neq 0 \text{ in } R, \text{ then } u \neq 0 \text{ in } R. \end{aligned} \tag{1}$$

For a given $u(p)$, let $u_1(p)$ be a solution of $L(u_1) = 0$, $u_1 - u \in S$. Then there exists a function $w(p)$, defined for $p \in R$, such that

$$u(p) = u_1(p) + v(p)\, L(u(w(p))). \tag{2}$$

Proof. Since $v(p)$ cannot vanish for $p \in R$, by virtue of the assumption of (1b), for every $p \in R$ there exists a number $a = a(p)$ such that

$$u(p) = u_1(p) + a v(p). \tag{3}$$

Since $[u(p) - u_1(p) - av(p)] \in S$, it follows from (1b) that we have $L[u(p) - u_1(p) - av(p)] = 0$ at some point in p. Call this point $w(p)$. Then, at this point,

$$0 = L[u(p) - u_1(p) - av(p)] = L(u) - a. \tag{4}$$

From this we see that $a = L(u(w(p)))$. This completes the proof.

Note that the burden of the proof of the mean-value theorem has been shifted to establishing the requisite positivity property. We shall discuss this point in some detail in § 21.

In their paper, HARTMAN and WINTNER presented a number of interesting examples associated with work of BLASCHKE [3], ZAREMBA [4], and others.

§ 20. Generalized Taylor Expansions

A generalized mean-value theorem is, of course, the first step toward a generalized Taylor expansion. Although it would take us too far off course to discuss these matters, we would like to refer the interested reader to the papers by DELSARTE [1], LEVITAN [2], PETERSSON [16.1], and WIDDER [3].

§ 21. Positivity of Operators

In the rest of the chapter, we shall devote our attention in the main to a survey of some techniques that can be used to establish the positivity of the classical operators of mathematical physics. We shall omit any discussion of direct verification by means of an explicit representation of the solution, omit the standard techniques of the theory of partial differential equations based on maximum principles (see NURENBERG [1] and WEINBERGER [2]), and, finally, omit the intuitive proofs based on the connection between GREEN's functions and stochastic processes (see KAC [1.9]).

We shall instead indicate the extension of the variational approach of § 8, as given by BELLMAN-BOCHNER [3], [4], and sketch a proof based on finite differences [5]. This last uses the idea of a random-walk process without explicitly mentioning the fact.

§ 22. Elliptic Equations

It is easily seen that the method presented in § 12 in connection with the second-order ordinary differential equation (12.1) may be applied to the multidimensional equation

$$\begin{aligned} &u_{xx} + u_{yy} + q(x, y)\, u = f(x, y)\,, \quad x, y \in R\,, \\ &u = 0\,, \quad x, y \in B\,, \end{aligned} \tag{1}$$

where B is the boundary of R.

Considering the quadratic functional $J(u)$, given by

$$J(u) = \int_R [u_x^2 + u_y^2 - q(x, y)\, u^2 + 2f(x, y)\, u]\, dx\, dy\,, \tag{2}$$

we readily derive the following result.

Theorem 15. *Let λ_1 be the smallest characteristic value of the Sturm-Liouville equation*

$$\begin{aligned} &u_{xx} + u_{yy} + \lambda u = 0\,, \quad x, y \in R\,, \\ &u = 0\,, \quad x, y \in B\,. \end{aligned} \tag{3}$$

If

$$q(x, y) \leqq \lambda_1 - d, \quad d > 0\,, \quad x, y \in R\,, \tag{4}$$

then the inequality

$$\begin{aligned} &u_{xx} + u_{yy} + q(x, y)\, u \geqq 0\,, \quad x, y \in R\,, \\ &u = 0\,, \quad x, y \in B\,, \end{aligned} \tag{5}$$

implies that

$$u \leqq 0\,, \quad x, y \in R\,. \tag{6}$$

The foregoing result is due to BELLMAN [12.1]. It is easy to see that it can be extended to Laplace-Beltrami operators for general domains.

§ 23. Positive Reproducing Kernels

A general theory of positive reproducing kernels that yields results concerning the positivity of GREEN's functions as special cases is due to ARONSZAJN and SMITH ([16.2] of Chapter 3). An interesting history of this problem and further references will be found there.

§ 24. Monotonicity of Mean Values

The following result, due to BELLMAN [21.3], in one sense belongs more properly in the following Chapter 5. As we shall see, however, following BOCHNER [21.4] in his discussion of quasi-analytic functions, it can be used to establish the positivity properties for parabolic operators. Let us begin with the simplest case.

Theorem 16. *If*

$$\begin{aligned} &u_t = u_{xx}, \quad 0 < x < 1, \quad t > 0, \\ &u(0, t) = u(1, t) = 0, \quad t > 0, \\ &u(x, 0) = g(x), \quad 0 < x < 1, \end{aligned} \tag{1}$$

then

$$I_n(t) = \int_0^1 [u(x, t)]^{2n} \, dx, \quad n = 1, 2, \ldots, \tag{2}$$

is a monotone decreasing function of t for $t \geqq 0$.

From this it follows that

$$\max_{0 \leqq x \leqq 1} |u(x, t)| \tag{3}$$

is a monotone decreasing function of t for $t \geqq 0$.

The result can easily be extended to cover general types of parabolic equations, as will be seen from the proof (cf. also BOCHNER's paper [21.4], cited above). The result that the function (3) is monotone decreasing had previously been obtained by PÓLYA and SZEGÖ [1], who used a different technique.

Proof. We have

$$\begin{aligned} \frac{dI_n}{dt} &= 2n \int_0^1 [u(x, t)]^{2n-1} u_t \, dx \\ &= 2n \int_0^1 [u(x, t)]^{2n-1} u_{xx} \, dx \\ &= 2n \{[u(x, t)]^{2n-1} u_x\}_0^1 - 2n(2n-1) \int_0^1 [u(x, t)]^{2n-2} u_x^2 \, dx \\ &= -2n(2n-1) \int_0^1 u^{2n-2} u_x^2 \, dx . \end{aligned} \tag{4}$$

Hence, $dI_n/dt < 0$.

To show that the function (3) has the desired property, we use the following lemma, of interest in itself.[1]

Lemma. *If $v(x)$ is a continuous function in a finite interval $[a, b]$, then*

$$\lim_{n\to\infty}\left\{\int_a^b [v(x)]^{2n}\,dx\right\}^{1/2n} = \max_{a\le x\le b} |v(x)|\,. \tag{5}$$

The proof is easily obtained. Let $m = \max |v(x)|$ in $[a, b]$; then

$$m^{2n}(b-a) \geqq \int_a^b [v(x)]^{2n}\,dx \geqq \int_c^d [v(x)]^{2n}\,dx\,, \tag{6}$$

where $[c, d]$ is a subinterval of $[a, b]$ within which $|v(x)| \geqq m-\varepsilon$. Thus

$$m\,(b-a)^{1/2n} \geqq \left\{\int_a^b [v(x)]^{2n}\,dx\right\}^{1/2n} \geqq (m-\varepsilon)\,(d-c)^{1/2n}\,. \tag{7}$$

Since

$$\lim_{n\to\infty} q^{1/2n} = 1\,, \tag{8}$$

for any $q > 0$, we see that (5) holds.

Combining the two results, we have a proof of the theorem. The preceding lemma appears first to have been used by M. Riesz.

§ 25. Positivity of the Parabolic Operator

Let us now, following Bochner [21.4], use the monotone behavior of the function

$$\max_{0\le x\le 1} |u(x, t)|\,,$$

established above, to show the positivity of the parabolic operator with appropriate boundary conditions.

The solution of

$$u_t = u_{xx},\quad u(x, 0) = f(x)\,, \tag{1}$$

$t > 0$, $f(x)$ periodic in x of period 2π, is given by

$$u(x, t) = \frac{1}{2\pi}\int_0^{2\pi} G(x-y, t)\, f(y)\,dy\,, \tag{2}$$

where

$$G(x, t) = 1 + 2\sum_{r=1}^{\infty} e^{-r^2 t}\cos r x\,. \tag{3}$$

[1] This is in the same spirit as the quasi linearization used in Chapter 1. A functional of complicated structure is written as a limit of functionals of simpler type.

The positivity of $G(x, t)$ follows directly from the functional equation for the theta function, namely, the fundamental transformation formula

$$G(x, t) = \left(\frac{\pi}{t}\right)^{1/2} \sum_{r=-\infty}^{\infty} \exp \frac{-(x - 2\pi r)^2}{4t} . \tag{4}$$

Relations of the type (4), however, are not available for general regions, whereas the following argument is independent of the region.

Suppose that $G(x, t)$ were negative for some $t > 0$ and some x. If this value of t is kept fixed, the normalization condition

$$\frac{1}{2\pi} \int_0^{2\pi} G(y, t)\, dy = 1 \tag{5}$$

implies that there must be an open set R such that

$$\frac{1}{2\pi} \int_R G(y, t)\, dy > 1 . \tag{6}$$

Hence, there is an interval $I \subset R$ such that

$$\frac{1}{2\pi} \int_I G(y, t)\, dy > 1 . \tag{7}$$

We now proceed along classic lines in the theory of partial differential equations. It is possible to construct a continuous smoothing function $f(x)$ such that

$$\begin{aligned} &\text{(a)} \quad 0 \leqq f(x) \leqq 1 \quad \text{for all } x, \\ &\text{(b)} \quad f(x) = 1 , \quad x \in I , \\ &\text{(c)} \quad f(x) = 0 \quad \text{outside } R. \end{aligned} \tag{8}$$

For this function, and for all x and t, we have

$$|u(x, t)| \leqq \max_x |f(x)| = 1; \tag{9}$$

but for the value of t for which $G(x, t)$ is assumed to be negative, we have

$$\begin{aligned} u(0, t) &= \frac{1}{2\pi} \int_0^{2\pi} G(y, t) f(y)\, dt \\ &= \frac{1}{2\pi} \int_I G(y, t)\, dy > 1 , \end{aligned} \tag{10}$$

a contradiction.

§ 26. Finite-difference Schemes

Consider, as an approximation to the solution of the parabolic equation (25.1), the finite-difference scheme

$$u(x, t + \delta^2) = \frac{u(x + \delta, t) + u(x - \delta, t)}{2} , \tag{1}$$

defined over the grid

$$
\begin{aligned}
x &= 0, \delta, \ldots, n\delta = 1\,, \\
t &= 0, \delta^2, \ldots,
\end{aligned}
\tag{2}
$$

with the boundary values

$$
\begin{aligned}
&\text{(a)} \quad u(0, t) = u(1, t) = 0\,, \\
&\text{(b)} \quad u(x, 0) = f(x)\,,
\end{aligned}
\tag{3}
$$

holding over the set of discrete x-values.

It is clear that the solution of the recurrence relation is nonnegative if $f(x) \geqq 0$. Since the limit of the recurrence relation as $\delta \to 0$ is the partial differential equation, it is plausible that the limit of the solution of the recurrence relation is the solution of the partial differential equation.

There are several ways that we can proceed. We can prove that the statement above is valid under reasonable conditions on $f(x)$ *without* assuming the existence of a solution of the partial differential equation (see JOHN [1]); or we can prove that it is valid on the assumption that the solution does exist — a much simpler result; or we can show that the kernel function for the discrete case approaches the kernel function for the continuous case as $\delta \to 0$. This last approach is not difficult to carry out for the case of constant coefficients.

Using the finite-difference algorithm,

$$
\begin{aligned}
u(x, t + \delta^2) = {} & \frac{u(x + \delta, t) + u(x - \delta, t)}{2} \\
& + \int_{x - q\delta}^{x + q\delta} u(y, t)\, dy\,,
\end{aligned}
\tag{4}
$$

we can establish the following result on the assumption that we have already established the existence of a solution of the partial differential equation.

Theorem 17. *If*

$$
\begin{aligned}
&\text{(a)} \quad u_t - u_{xx} + q(x, t)\, u \geqq 0\,, \quad 0 < x < 1\,, \quad t > 0\,, \\
&\text{(b)} \quad u(x, 0) \geqq 0\,, \quad 0 \leqq x \leqq 1\,, \\
&\text{(c)} \quad u(0, t) = u(1, t) = 0\,, \quad t > 0\,, \\
&\text{(d)} \quad q(x, t) \geqq -k(T) > -\infty\,, \quad 0 \leqq x \leqq 1\,, \quad 0 \leqq t \leqq T, \\
&\qquad \textit{for any } T > 0,
\end{aligned}
\tag{5}
$$

then $u(x, t) \geqq 0,\ 0 < x < 1,\ t > 0.$

The proof is easily carried through, once we note that a transformation of the form $u = e^{\lambda t} y$ permits us to assume that $q(x, t) \geqq 0$.

The point of the formula (4) is that by choice of the *appropriate* difference relation we can make nonnegativity of the solution apparent. We shall discuss this method again in § 27, below.

§ 27. Potential Equations

The technique of § 26 can be used to treat the potential equation

$$\begin{aligned} &u_{xx} + u_{yy} = f(x, y)\,, \quad (x, y) \in R\,, \\ &u = 0\,, \quad (x, y) \in B\,. \end{aligned} \tag{1}$$

Consider the recurrence relation

$$\begin{aligned} u(x, y) = \frac{u(x+\delta, y) + u(x-\delta, y) + u(x, y+\delta) + u(x, y-\delta)}{4} - \\ - 2f(x, y)\,\delta^2, \end{aligned} \tag{2}$$

defined over an (x, y)-grid, $x, y = \pm k\delta$, which reduces formally to (1) as $\delta \to 0$.

In place of this *static* recurrence relation, which does not render the positivity property at all obvious, consider the *dynamic relation*

$$\begin{aligned} u_{n+1}(x, y) = \frac{u_n(x+\delta, y) + u_n(x-\delta, y) + u_n(x, y+\delta) + u_n(x, y-\delta)}{4} - \\ - 2f(x, y)\,\delta^2. \end{aligned} \tag{3}$$

If $u_n(x, y) \leqq 0$ and $f(x, y) \geqq 0$, it is clear that $u_{n+1}(x, y) \leqq 0$.

In this way we can establish the fact that the inequality

$$\begin{aligned} &u_{xx} + u_{yy} \geqq 0\,, \quad (x, y) \in R\,, \\ &u = 0\,, \quad (x, y) \in B\,, \end{aligned} \tag{4}$$

implies that $u \leqq 0$ under appropriate assumptions concerning the boundary B, again easily on the assumption that we have by other means established the existence of a solution of the partial differential equation, or by more difficult arguments if we wish to start *ab initio*.

§ 28. Discussion

Once again, we wish to emphasize the fact that there are often many discrete versions of the same continuous equation, with the obvious consequence that some are better suited than others for analytic or computational purposes.

As a further example, consider the nonlinear equation

$$u_t = u u_x, \quad u(x, 0) = g(x)\,. \tag{1}$$

In place of one of the usual schemes, we can employ the difference equation

$$u(x, t+\delta) = u(x + \delta u(x, t), t)\,, \tag{2}$$

which renders the nonnegativity of the solution, and its uniform boundedness, apparent.

Similarly, one can treat the equation of BURGERS,

$$u_t = u u_x + \varepsilon u_{xx}\,, \tag{3}$$

by the same method. The relation (2) has been used for computationale purposes with considerable success [1].

§ 29. The Inequalities of Haar-Westphal-Prodi

The results of §§ 24 to 26 concerning the linear parabolic equation are special cases of more general theorems connected with the nonlinear inequality

$$u_t < g\,(u_{xx}, u_x, u, x, t)\,. \tag{1}$$

This study was initiated by HAAR [1] and continued by WESTPHAL [2], PRODI [3], PUCCI [4], and MLAK [5]. The resultant inequalities play an important role in the study of the stability of the solutions of nonlinear parabolic equations; see BELLMAN [6], NARASIMHAN [7], and MCNABB [8].

§ 30. Some Inequalities of Wendroff

The results of § 4 can be extended in a number of ways. Following are some unpublished inequalities due to WENDROFF:

If

$$u\,(x, y) \leqq c + \int_0^y \int_0^x v\,(r, s)\, u\,(r, s)\, dr\, ds\,, \tag{1}$$

where $c \geqq 0$, $u\,(r, s)$, $v\,(r, s) \geqq 0$, then

$$u \leqq c\, e^{\int_0^y \int_0^x v(r,s)\,dr\,ds}\,. \tag{2}$$

If

$$u\,(x, y) \leqq a(x) + b(y) + \int_0^y \int_0^x v\,(r, s)\, u\,(r, s)\, dr\, ds\,, \tag{3}$$

where $a(x)$, $b(y) > 0$, $a'(x)$, $b'(y) \geqq 0$, $u, v \geqq 0$, then

$$u \leqq \frac{[a(0) + b(y)]\,[a(x) + b(0)]\, e^{\int_0^y \int_0^x v\,dr\,ds}}{a(0) + b(0)}\,. \tag{4}$$

If

$$u\,(x, y) \leqq c + a \int_0^x u\,(x, y)\, ds + b \int_0^y u\,(x, s)\, ds\,, \tag{5}$$

then

$$u\,(x, y) \leqq c\, e^{ax + by + ab\,xy}\,. \tag{6}$$

If

$$u(x, y) \leqq a(x) + b(y) + a\int_0^x u(s, y)\, ds + b\int_0^y u(x, s)\, ds\,, \tag{7}$$

then

$$u(x, y) \leqq Q(x, y), \tag{8}$$

where $Q(x, y)$ denotes the function

$$\frac{\left[a(0) + b(0) + \int_0^y e^{-b y_1} b'(y_1)\, dy_1\right]\left[[a(0) + b(0) + \int_0^x e^{-a x_1} a'(x_1)\, dx_1\right] e^{ax+by+abxy}}{[a(0) + b(0)]}$$

§ 31. Results of Weinberger-Bochner

It was shown by WEINBERGER [21.2] that some results of BOCHNER [1], [3], which in turn are generalization of earlier results of FEJÉR, concerning nonnegativity properties of trigonometric polynomials could be interpreted in terms of positivity properties of solutions of hyperbolic equations and the associated Riemann function. These results tie together in a very interesting fashion the positivity results for polynomials and trigonometric polynomials mentioned in §§ 10 and 11 of Chapter 3 and the positivity properties of linear differential operators.

§ 32. Variation-diminishing Transformations

A natural extension of the concept of a positive transformation is that of a variation-diminishing transformation. By this we mean a transformation of the type

$$v(x) = \int_a^b k(x, y)\, u(y)\, dy\,, \tag{1}$$

possessing the property that the number of changes of sign of $v(x)$ in the interval $[a, b]$ is less than or equal to the number of changes of sign of $u(y)$ in this interval.

An interesting and comprehensive expository discussion of various problems arising in this way may be found in SCHOENBERG [1.10].

§ 33. Quasi Linearization

One reason for our interest in the nonnegativity of solutions of linear inequalities lies in the fact that results of this nature can be used in obtaining representations of solutions of nonlinear differential equations.

Consider, for example, the Riccati differential equation

$$\frac{du}{dt} = u^2 + a(t)\,, \quad u(0) = c\,. \tag{1}$$

Since

$$u^2 = \max_v\,(2uv - v^2)\,, \tag{2}$$

a particular case of the quasi-linear representation of convex functions referred to in § 26 of Chapter 1, we may write (1) in the form

$$\frac{du}{dt} = \max_v \left[2uv - v^2 + a(t)\right], \quad u(0) = c. \tag{3}$$

Hence for any function $v(t)$, we have the inequality

$$\frac{du}{dt} \leqq 2uv - v^2 + a(t), \quad u(0) = c. \tag{4}$$

Referring to Theorem 1 of § 2, we see that this means that $u \geqq U$, where U is the solution of the equation

$$\frac{dU}{dt} = 2Uv - v^2 + a(t), \quad U(0) = c. \tag{5}$$

Since U may be written

$$U = e^{2\int_0^t v\,ds} c + \int_0^t [a(t_1) - v^2]\, e^{2\int_{t_1}^t v\,ds} dt_1, \tag{6}$$

we see that we can express u, the solution of (1), in the form

$$u = \max_v \left[e^{2\int_0^t v\,ds} + \int_0^t [a(t_1) - v^2]\, e^{2\int_{t_1}^t v\,ds} dt_1 \right]. \tag{7}$$

The maximum is attained for $v = u$.

A further discussion of these matters may be found in BELLMAN [1] and COLLATZ [2]. See also [26.1], [26.2], and [26.3] of Chapter 1, where a number of other types of functional equations are discussed.

§ 34. Stability of Operators

The questions we study in this chapter may be considered to be particular cases of a still more general class of problems concerning the stability of functional transformations.

Given an operator T and a solution class $\{u\}$ with the property that $T(u) = 0$, when does $\|T(v)\| \leqq \varepsilon$ (where $\|\ \|$ denotes some suitable norm) imply that $\|v - u\| \leqq \delta(\varepsilon)$ for some u?

For example, if $f(x)$ is a real continuous function of x over $(-\infty, \infty)$, and

$$|f(x+y) - f(x) - f(y)| \leqq \varepsilon, \tag{1}$$

it was shown by HYERS and ULAM [1] that there exists a constant k such that

$$|f(x) - kx| \leqq 2\varepsilon. \tag{2}$$

For some results pertaining to approximately convex functions, see GREEN [2]. Although a great deal of work in this direction has been done in connection with ordinary and partial differential equations, little has been done for other types of functional equations. For the classical results, see the books by CODDINGTON-LEVINSON [3], LEFSCHETZ [4], CESARI ([19.4] of Chapter 3), and BELLMAN [1.3].

§ 35. Miscellaneous Results

The subject of functional inequalities is full of isolated results, indicative of the existence of general theories, but as yet incompletely related.

There are many connections with Tauberian theory; see WIDDER ([1.8] of Chapter 3), DOETSCH [1], and the paper by WRIGHT [2], in which a differential inequality leading to the prime-number theorem along the ERDÖS-SELBERG path is discussed. See also SHAPIRO [3].

A most interesting result, with a number of immediate applications, is the following: If $u_m \geqq 0$, and

$$u_{m+n} \leqq u_m + u_n \,, \tag{1}$$

for $m, n \geqq 0$, then

$$\lim_{n \to \infty} \frac{u_n}{n} \tag{2}$$

exists. For a proof, see PÓLYA-SZEGÖ ([44.1] of Chapter 1); for applications, see FURSTENBERG and KESTEN [4], and BELLMAN [5].

Further developments, by BECKENBACH and RADÓ [17.15], [17.16], BECKENBACH [17.17], CALABI [6], YANO and BOCHNER [7], OSSERMAN [8], REDHEFFER [9], [14], SACKSTEDER [10], BOAS and PÓLYA [11], DUFFIN and SERBYN [12], PAYNE and WEINBERGER [13], and LOEWNER [47.2] of Chapter 2, are indicated in the bibliography.

Bibliographical Notes

§ 1. The interesting contributions of Čaplygin concerning the positivity of operators are most readily found in the monograph

1. ČAPLYGIN, S. A.: New methods in the approximate integration of differential equations (Russian). Moscow: Gosudarstv. Izdat. Tech.-Teoret. Lit. 1950.
2. GRONWALL, T. H.: Note on the derivatives with respect to a parameter of the solutions of a system of differential equations. Ann. of Math. **20**, 292—296 (1918).

For a discussion of stability theory and its connection with inequalities, see

3. BELLMAN, R.: Stability theory of differential equations. New York: McGraw-Hill Book Co., Inc. 1954.
4. BIHARI, I.: A generalization of a lemma of BELLMAN and its application to uniqueness problems of differential equations. Acta Math. Hung. **7**, 81—94 (1956).
5. LANGENHOP, C. E.: Bounds on the norm of a solution of a general differential equation. Proc. Am. Math. Soc. **11**, 795—799 (1960).

The fundamental results of Pólya were stimulated by a note of Poincaré:

6. Poincaré, H.: L'Intermediaire des Mathématiciens **1**, 141—144 (1894).

See also the comments of H. Dellac, J. Hadamard, J. Roux and E. Duporcq on pp. 69—70, 127, 172—173, and 216—217 of this same volume of L'Intermediare des Mathématiciens.

7. Pólya, G.: On the mean-value theorem corresponding to a given linear homogeneous differential equation. Trans. Am. Math. Soc. **24**, 312—324 (1922).
8. Hartman, P., and A. Wintner: Mean value theorems and linear operators. Am. Math. Monthly **62**, 217—222 (1955).

In our brief study of inequalities involving ordinary differential and partial differential operators, we face the usual difficulty of the study of inequalities: It appears to be very difficult to draw the line between what should be included in the theory of inequalities and what should be relegated to the theory of differential equations.

9. Kac, M.: On some connections between probability theory and differential and integral equations. Proc. Second Berkeley Symposium on Mathematical Statistics and Probability. Berkeley and Los Angeles, Calif.: University of California Press 1951.
10. Schoenberg, I. J.: On variation-diminishing transformations. Proc. Symposium on Numerical Approximation, 1958. Madison, Wis.: University of Wisconsin 1959.

§ 4. This result, and its application to stability theory, was first given in

1. Bellman, R.: The stability of solutions of linear differential equations. Duke Math. J. **10**, 643—647 (1943).

For other applications, see

2. Nemyckii, V. V., and V. V. Stepanov: Qualitative theory of differential equations (Russian). Moscow: OGIZ 1947.
3. Guiliano, L.: Generalazzioni di un lemma di Gronwall. Rend. Accad. Lincei **1946**, 1264—1271.

For an application of this inequality to the derivation of a priori estimates for solutions of partial differential equations, see

4. Lax, P. D.: The initial value problem for nonlinear hyperbolic equations in two independent variables. Chapter 12 of Annals of Math. Studies, No. 33. Princeton, N. J.: Princeton University Press 1954.

See also

5. Lax, P. D.: The scope of the energy method. Bull. Am. Math. Soc. **66**, 32—35 (1960),
6. Lax, P. D.: On Cauchy's problem for hyperbolic equations and the differentiability of solutions of elliptic equations. Comm. Pure Appl. Math. **8**, 615—633 (1955),

where many other references will be found. Differential inequalities also play a role in the application of Liapunov's second method to the establishment of stability; see

7. LaSalle, J. P., and S. Lefschetz: The stability theory of Liapunov. New York: Academic Press 1961.
8. Lees, M.: Appoximate solutions of parabolic equations. J. Soc. Ind. Appl. Math. **7**, 167—183 (1959).

§ 6. This result is given in [43.1] of Chapter 2. The physical significance of the condition is revealed in the study of the multistage production processes mentioned in Chapter 3, and also in the study of Markovian decision processes; cf. [7.1] of Chapter 1, [44.2] of Chapter 2, and

1. Romanovsky, E. R.: On a theorem of R. Bellman. Theory of probability and its applications **4**, 456—458 (1959).

2. OPIAL, Z.: Sur un système d'inégalités intégrales. Ann. Polon. Math. 3, 200—289 (1957).

§ 7. This proof by O. TAUSSKY was communicated in a letter.

§ 8.

1. LOEWNER, C.: A theorem on the partial order derived from a certain transformation semigroup. Math. Z. **72**, 53—60 (1959).
2. GARNER, J. B., and L. P. BURTON: Solutions of linear differential systems satisfying boundary conditions in the large. Proc. Am. Math. Soc. **12**, 100—106 (1961).

§ 9.

1. PERRON, O.: Die Stabilitätsfrage bei Differentialgleichungen. Math. Z. **32**, 703—728 (1930).
2. MASSERA, J., and J. J. SCHÄFFER: Linear differential equations and functional analysis. I. Ann. of Math. **67**, 517—573 (1958).
3. MASSERA, J., and J. J. SCHÄFFER: Linear differential equations and functional analysis. II. Equations with periodic coefficients. Ann. of Math. **69**, 88—104 (1959).

PERRON's work was motivated by earlier work by BOREL and HARDY in this area.

§ 10.

1. FELLER, W.: On generalized Sturm-Liouville operators. Proc. Conference on Differential Equations. College Park, Md.: University of Maryland Press **1955**.

The results of FELLER given in this paper and in other works cited there are also closely connected with the generalized convexity we shall discuss below.

§ 12. This argument is given in

1. BELLMAN, R.: On the nonnegativity of GREEN's functions. Boll. Un. Math. **12**, 411—413 (1957).

It is closely related to a method used by STIELTJES,

2. STIELTJES, T. J.: Sur les racines de $X_n = 0$. Acta Math. **9**, 385—400 (1886—1887).

Extensive generalizations of this method, by ARROW, WONG, WOODBURY, and others, may be found in the references to Leontieff-Minkowski matrices given in § 38 of Chapter 2, and in [2.1] of that chapter.

Similarly, one can use the variational formulation to establish the variation-diminishing property of the GREEN's function. See

3. BELLMAN, R.: On the variation-diminishing property of GREEN's functions. Boll. Un. Math., to appear.

§ 15. An enormous amount of work has been done in this field. Let us merely cite

1. PETROV, V. N.: The limits of applicability of S. TCHAPLYGIN's theorem on differential inequalities to linear equations with usual derivatives of the second order. C. R. (Doklady) Acad. Sci. URSS **51**, 255—258 (1946),
2. WILKINS, J. E., JR.: The converse of a theorem of TCHAPLYGIN on differential equations. Bull. Am. Math. Soc. **53**, 126—129 (1947),

and the references given in § 18.

The decomposition theorem using the Wronskians enables us to see clearly the connection between positivity of a particular solution, characteristic values, and positivity of the operator in the case of a second-order linear differential operator. For higher-order equations, these connections have not been explored. Most likely, there is an intimate relation with the theory of variation-diminishing transformations, as mentioned again in the subsequent § 16.

§ 16.

1. PETERSSON, H.: Über Interpolation durch Losungen linearer Differentialgleichungen. Abh. Math. Sem. Univ. Hamburg **16**, 41—55 (1949).

§ 17.

1. BECKENBACH, E. F.: Generalized convex functions. Bull. Am. Math. Soc. **43**, 363—371 (1937).
2. BECKENBACH, E. F., and R. H. BING: On generalized convex functions. Trans. Am. Math. Soc. **58**, 220—230 (1945).
3. VALIRON, G.: Fonctions convexes et fonctions entières. Bull. Soc. Math. France **60**, 278—287 (1932).
4. BONSALL, F. F.: The characterization of generalized convex functions. Quart. J. Math. **1**, 100—111 (1950).
5. PEIXOTO, M.: Convexidas das curvas. Notas de matematica, No. 6. Livraria Boffoni 1948.
6. PEIXOTO, M.: On the existence of a derivative of generalized convex functions. Summa Brasiliensis Math. **2**, No. 3 (1948).
7. PEIXOTO, M.: Generalized convex functions and second-order differential inequalities. Bull. Am. Math. Soc. **55**, 563—572 (1949).

See also

8. MOTZKIN, T. S.: Approximation by curves of a unisolvent family. Bull. Am. Math. Soc. **55**, 789—793 (1949).
9. TORNHEIM, L.: On n-parameter families of functions and associated convex functions. Trans. Am. Math. Soc. **69**, 457—467 (1950).
10. CURTIS, P. C., JR.: N-parameter families and best approximation. Pacific J. Math. **9**, 1013—1027 (1959).
11. REID, W. T.: Variational aspects of generalized convex functions. Pacific J. Math. **9**, 571—581 (1954).
12. HARTMAN, P.: Unrestricted n-parameter families. Rend. Circ. Mat. Palermo (2) **7**, 123—142 (1958).

These ideas carried over to the realm of partial differential operators lead naturally to subharmonic functions and their generalizations; see

13. RIESZ, F.: Sur les fonctions subharmoniques et leur rapport á la théorie du potential, I, II. Acta Math. **48**, 329—343 (1926); **54**, 321—360 (1930).
14. RADÓ, T.: Subharmonic functions. Ergeb. Math. Berlin: J. Springer Verlag 1937.
15. BECKENBACH, E. F., and T. RADÓ: Subharmonic functions and minimal surfaces. Trans. Am. Math. Soc. **35**, 648—661 (1933).
16. BECKENBACH, E. F., and T. RADÓ: Subharmonic functions and surfaces of negative curvature. Trans. Am. Math. Soc. **35**, 662—674 (1933).
17. BECKENBACH, E. F.: On subharmonic functions. Duke Math. J. **1**, 481—483 (1935).
18. READE, M.: Some remarks on subharmonic functions. Duke Math. J. **10**, 531—536 (1943).
19. BONSALL, F. F.: On generalized subharmonic functions. Proc. Cambridge Phil. Soc. **46**, 387—395 (1950).
20. TAUTZ, G.: Zur Theorie der elliptischen Differentialgleichungen. I. Math. Ann. **117**, 694—726 (1939—1941).
21. BECKENBACH, E. F., and L. K. JACKSON: Subfunctions of several variables. Pacific J. Math. **3**, 291—313 (1953).
22. JACKSON, L. K.: On generalized subharmonic functions. Pacific J. Math. **5**, 215—228 (1955).
23. INONE, M.: Dirichlet problem relative to a family of functions. J. Inst. Polytech. Osaka City University (A) **7**, 1—16 (1956).
24. JACKSON, L. K.: Subfunctions and the Dirichlet problem. Pacific J. Math. **8**, 243—255 (1958).

§ 18.

1. KLIMKO, E. Y.: Some examples of the method of CHAPLYGIN in the approximation of solutions of differential equations with retarded arguments. Uspekhy Mat. Nauk **12**, 305—312 (1957).
2. ARTEMOV, G. A.: Application of ČAPLYGIN's method to the solution of the characteristic CAUCHY's problem for a partial differential equation of parabolic type. C. R. (Doklady) Akad. Sci. URSS **112**, 791—792 (1957).
3. PARODI, M.: Détermination de courbes planes définies par une inégalité entre les valeurs absolues de fonctions des éléments de contact et un point courant. C. R. Acad. Sci. (Paris) **245**, 1871 (1957); **246**, 3573—3574 (1958).

§ 19.

1. PÓLYA, G.: Quelques théorèmes analogues au théorème de ROLLE, lies a certaines equations lineaires aux derivees partielles. C. R. Acad. Sci. (Paris) **199**, 655 (1934).
2. BLEULER, K.: Über den Rolleschen Satz für den Operator $\Delta u + \lambda u$ und die damit zusammengängenden Eigenschaften der Greenschen Funktion. Geneva: Zürich Doctoral Thesis 1942.
3. BLASCHKE, W.: Mittelwertsatz der Potentialtheorie. Jber. Dtsch. Math. Ver. **27**, 157—160 (1918).
4. ZAREMBA, S.: Sur une propriété générale des fonctions harmoniques. Conférence de la Réunion Internationale des Mathématiciens, 171—176. Paris 1937.

We thus observe a connection between two important concepts: mean-value theorems and positivity.

§ 20.

1. DELSARTE, J.: Sur une extension de la formule de TAYLOR. J. Math. Pures Appl. (9) **17**, 213—231 (1938).
2. LEVITAN, B. M.: Estimation of the remainder term in the formula of Taylor-Delsarte. C. R. (Doklady) Akad. Sci. URSS **73**, 269—272 (1950).
3. WIDDER, D. V.: A generalization of TAYLOR's series. Trans. Am. Math. Soc. 30, 126—154 (1928).

§ 21.

1. NURENBERG, L.: A strong maximum principle for parabolic equations. Comm. Pure Appl. Math. **6**, 167—177 (1953).
2. WEINBERGER, H. F.: A maximum property of CAUCHY's problem. Ann. of Math. **64**, 505—513 (1956).
3. BELLMAN, R.: A property of summation kernels. Duke Math. J. **15**, 1013—1019 (1948).
4. BOCHNER, S.: Quasi-analytic functions, Laplace operators, and positive kernels. Ann. of Math. **51**, 68—91 (1950).

In this last paper, the device used by BELLMAN in connection with the equation $u_t = u_{xx} + u_{yy} + u_{zz}$ is extended to the general Beltrami operator and applied to the wave equation $u_{tt} = u_{xx} + u_{yy} + u_{zz}$. These results are intimately connected with the general problem of obtaining a priori bounds for various functionals associated with a linear-operator equation.

The technique of using appropriate finite-difference approximations to render immediately obvious certain properties of the solution of a differential or partial differential equation is a very powerful one that has not as yet been extensively exploited. It is of particular importance in connection with questions of numerical solution. See, for example,

5. BELLMAN, R.: On the nonnegativity of solutions of the heat equation. Boll. Un. Math. **12**, 520—523 (1957).

§ 24.
1. PÓLYA, G., and G. SZEGÖ: Sur quelques propriétés qualitatives de la propagation de la chaleur. C. R. Acad. Sci. (Paris) **192**, 1340—1342 (1931).

§ 26. For a comprehensive account of convergence of finite-difference schemes, see
1. JOHN, F.: On integration of parabolic equations by difference methods. I. Linear and quasi-linear equations for the infinite interval. Comm. Pure Appl. Math. **5**, 155—211 (1952).

§ 28. See
1. BELLMAN, R., I. CHERRY and G. M. WING: A note on the numerical integration of a class of nonlinear hyperbolic equations. Quart. Appl. Math. **16**, 181—183 (1958).

§ 29. The paper by WESTPHAL [2] contains a generalization of a result first given by HAAR concerning partial differential inequalities. These results were extended, in turn, by PRODI, and used by MLAK and PRODI for stability purposes. The original paper by HAAR is
1. HAAR, A.: Über Eindeutigkeit und Analyzität der Lösungen partieller Differentialgleichungen. Atti Congresso Intern. Bologna **3**, 5—10 (1928).
2. WESTPHAL, H.: Zur Abschätzung der Lösungen nichtlinearer parabolischer Differentialgleichungen. Math. Z. **51**, 690—695 (1949).
3. PRODI, G.: Questioni di stabilita per equazioni non lineari alle derivate parziali di tipo parabolico. Acad. Naz. Lincei (8) **10**, 365—370 (1952).
4. PUCCI, C.: Ordering and uniqueness of solutions of boundary problems for elliptic equations. Technical Note BN-112. College Park, Md.: University of Maryland Press 1957.
5. MLAK, W.: Differential inequalities of parabolic type. Ann. Polon. Math. **3**, 349—354 (1957).

For a discussion of the stability problem for partial differential equations and the applicability of results of this nature, see
6. BELLMAN, R.: On the existence and boundedness of solutions of nonlinear partial differential equations of parabolic type. Trans. Am. Math. Soc. **64**, 21—44 (1948).
7. NARASIMHAN, R.: On the asymptotic stability of solutions of parabolic differential equations. J. Rat. Mech. Analys. **3**, 303—313 (1954).

Results of this nature were obtained in 1950 by P. LAX in the course of some joint unpublished work on the stability of solutions of parabolic equations. See also
8. MCNABB, A.: Notes on criteria for the stability of steady-state solutions of parabolic equations. J. Math. Analys. Appl., to appear.

§ 30. These results are heretofore unpublished.

§ 31.
1. BOCHNER, S.: Positive zonal functions on spheres. Proc. Nat. Acad. Sci. USA **40**, 1141—1147 (1954).
2. BOCHNER, S.: Sturm-Liouville and heat equations whose eigenfunctions are ultraspherical polynomials of associated Bessel function. Proc. of the conference on differential equations (dedicated to A. WEINSTEIN) 23—48. College Park, Md.: University of Maryland Book Store 1956.
3. PROTTER, M. H.: A maximum principle for hyperbolic equations in the neighborhood of an initial line. Berkeley and Los Angeles, Calif.: University of California Press 1956.

The basic problem is that of investigating sets of function $\{f_n(x)\}$ possessing the property that

$$\sum_{n=1}^{\infty} a_n f_n(x) \geqq 0$$

in $-1 \leqq x \leqq 1$ implies

$$\sum_{n=1}^{\infty} a_n f_n(x) f_n(y) \geqq 0$$

for $-1 \leqq x \leqq 1, \quad -1 \leqq y \leqq 1$.

§ 32. See the references in § 48 of Chapter 2.

§ 33. A detailed discussion of the connection with NEWTON's method for functional equations, together with numerical examples, is given in [26.3] of Chapter 1. Other results are given in [26.1] and [26.2] of that chapter, and in

1. BELLMAN, R.: On monotone convergence to solutions of $u' = g(u, t)$. Proc. Am. Math. Soc. 8, 1007—1009 (1957).

Many further references and applications will be found in the paper by KALABA cited above ([26.3] of Chapter 1) and in

2. COLLATZ, L., in R. LANGER (editor): Applications of the theory of monotonic operators to boundary value problems. Boundary problems in differential equations. Madison, Wis.: University of Wisconsin 1960.

§ 34.

1. HYERS, D., and S. ULAM: Approximately convex functions. Proc. Bull. Am. Math. Soc. 3, 821—828 (1952).
2. GREEN, J. W.: Approximately convex functions. Duke Math. J. 19, 499—504 (1952).
3. CODDINGTON, E., and N. LEVINSON: Theory of ordinary differential equations. New York: McGraw-Hill Book Co., Inc. 1955.
4. LEFSCHETZ, S.: Differential equations: geometric theory. New York: Interscience Publishers, Inc. 1957.

§ 35.

1. DOETSCH, G.: Theorie und Anwendung der Laplace-Transformation. New York: Dover Publications 1943.
2. WRIGHT, E. M.: Functional inequalities in the elementary theory of primes. Duke Math. J. 19, 695—704 (1952).
3. SHAPIRO, H. N.: Tauberian theorems and elementary prime number theory. Comm. Pure Appl. Math. 12, 579—610 (1959).
4. FURSTENBERG, H., and H. KESTEN: Products of random matrices. Ann. Math. Stat. 31, 457—469 (1960).
5. BELLMAN, R.: Functional equations on the theory of dynamic programming. Rend. Circ. Mat. Palermo (2) 8, 1—3 (1959).

We have avoided any discussion of a large number of interesting results concerning partial differential inequalities of the form $\Delta u \geqq f(u)$ and their applications to differential geometry and analysis. See, for example, [17.15], [17.16], and the following:

6. CALABI, E.: An extension of E. HOPF's maximum principle with an application to Riemannian geometry. Duke Math. J. 25, 45—46 (1958),
7. YANO, K., and S. BOCHNER: Curvature and Betti numbers. Ann. Math. Stud., No. 32. Princeton, N. J.: Princeton University Press 1953,
8. OSSERMAN, R.: On the inequality $\Delta u \geqq f(u)$. Pacific J. Math. 7, 1641—1647 (1957),

where references to earlier work by WITTICH, HAVILAND, and WALTER may be found.

9. REDHEFFER, R.: On the inequality $\Delta u \geqq f(u, |\mathrm{grad}\, u|)$. J. Math. Analys. Appl. 1, 277—299 (1960).
10. SACKSTEDER, R.: On local and global properties of convex sets and hypersurfaces, AFOSR TN 59-1294. Baltimore, Md.: Johns Hopkins University Press 1959.

In this chapter we have completely omitted all aspects of completely convex functions and the important results of BERNSTEIN-HAUSDORFF-WIDDER and generalizations. See [1.8] of Chapter 3 and

11. BOAS, R. P., and G. PÓLYA: Generalizations of completely convex functions. Proc. Nat. Acad. Sci. USA **27**, 323—325 (1941).

Let us also note that we have not referred to a variety of techniques based on variational methods. See

12. DUFFIN, R. J., and W. D. SERBYN: Approximate solution of differential equations by a variational method. J. Math. Physics **37**, 162—168 (1958).
13. PAYNE, L. E., and H. F. WEINBERGER: New bounds for the solutions of second order elliptic partial differential equations. College Park, Md.: University of Maryland, Tech. Note BN-108, 1959.

Finally, let us refer to

14. REDHEFFER, R.: Inequalities for a matrix Riccati equation. J. Math. Mech. **8**, 349—367 (1959),

and again to [47.2] in Chapter 2, a paper first cited in connection with monotone matrix functions.

Chapter 5

Inequalities for Differential Operators

§ 1. Introduction

In this concluding chapter, we shall pursue some variations on yet another central theme in analysis. The present theme can be described in the following general terms. Consider a set of ordinary differential operators, $\{T_i\}$, and a set of functions, $\{u\}$, admitting these operators. These give rise to a new set of functions $\{T_i u\}$. We have already considered the problem of determining when nonnegativity of the $T_i u$ induces a corresponding property in the u. In this chapter, we shall consider the problem of determining when the fact that $T_i u \in L^p(0, \infty)$ implies that $u \in L^r(0, \infty)$, where the value of r depends, of course, on p. More generally, given that $T_i u \in L^{p_i}(0, \infty)$, $i = 1, 2, \ldots, k$, we wish to determine the L-class of $T_{k+1} u$.

This problem, in turn, is a particular case of the problem of obtaining a complete set of inequalities of the form

$$\int_0^\infty |T_{k+1} u|^r \, dt \leq g\left(\int_0^\infty |T_1 u|^{p_1} \, dt, \ldots, \int_0^\infty |T_k u|^{p_k} \, dt\right). \tag{1}$$

We shall not discuss this type of question, although we shall indicate some of the inequalities of this nature that can be obtained.

Perhaps the most famous inequality of the foregoing type is the classical one due to S. BERNSTEIN: If

$$u = \sum_{n=0}^{N} a_n e^{int},$$

a trigonometric polynomial of degree N, then

$$\max_{0 \leqq t \leqq 2\pi} |u'(t)| \leqq N \max_{0 \leqq t \leqq 2\pi} |u(t)| \,. \tag{2}$$

A corresponding result, due to MARKOFF, holds for ordinary polynomials over a finite interval; see MANDELBROJT [1].

There are numerous extensions of these results — to other norms, to generalized trigonometric polynomials, to entire functions of finite order, and so on. We shall not enter into any of these matters here, since they have been thoroughly and elegantly treated in the book by BOAS [2] and in the expository paper by SCHAEFFER [3]. A host of references to the enormous work in this field will be found in these two soürces. See also SZ.-NAGY [4].

As in the other parts of this monograph, our aim is to focus attention upon results and methods that have not heretofore been collected or analyzed in any detail. Furthermore, it is our aim to present particular methods that can be used in a variety of ways.

Also as in the other parts of this monograph, it has been difficult to draw a sharp distinction between inequalities of general interest throughout analysis, and those useful only in specialized fields. Despite their great elegance and importance, we have somewhat relucantly decided to omit any of the numerous inequalities concerning partial differential operators. A discussion of the inequalities of POINCARÉ, KORN, FRIEDRICHS, and others, may be found in the papers by ARONSZAJN [5], NIRENBERG [6], and FRIEDRICHS [7]; see also COURANT-HILBERT ([26.1] of Chapter 2). We feel that these results lie most firmly imbedded in the domain of partial differential equations.

The results we shall begin with have their inception in the investigation of HADAMARD concerning relations that exist between bounds for $u(t)$, $u'(t)$, and $u''(t)$ for t in a finite interval. A discussion of this is contained in HARDY, LITTLEWOOD, and PÓLYA ([1.1] of Chapter 1). For an indication of how results of this nature are connected with Tauberian theory, see the book by TITCHMARSH [8].

Generalizations of these results were given by ESCLANGON and LANDAU [9]. The problem of determining best possible inequalities connecting various norms of u, u', and u'' was undertaken by KOLMOGOROFF [10]. As mentioned above, we shall not discuss any questions of this type.

Problems of the sort we are considering in this chapter are intimately connected with the theory of linear differential equations, as has been noted by a number of authors (BEESACK [11], [12], COLAUTTI [13], REID [14], BELLMAN [15], and others), and we shall exploit this connection.

Following a treatment of results concerning L^p-norms, due to NAGY [16], HALPERIN and VON NEUMANN [17], HALPERIN and PITT [18], and BELLMAN [15], we shall turn to a class of inequalities associated with the name of CARLSON [19]. As pointed out by NAGY [16], these are related to the Hadamard-type inequalities; and, as pointed out by KJELLBERG [20], they are also related to the moment problems discussed in Chapter 3.

Extensions of CARLSON's inequality were given by HARDY [21], GABRIEL [22], BEURLING [23], CATON [24], NAGY [16], KJELLBERG [20], LEVIN [26], and BELLMAN [25]. Without paying any attention to the determination of best possible constants, a problem solved by LEVIN [26], we shall follow a method that furnishes a simple means of obtaining results of this nature.

Next we turn to WIRTINGER's inequality, a result used in the preceding Chapter 4, and to its numerous extensions. We shall employ an interesting method based on explicit identities to establish these results. Although they can readily be obtained by means of the theory of Sturm-Liouville equations, it is important to use this elementary method, due to BEESACK [12], since it covers a number of cases that would otherwise require an extensive background in analysis. The method was highly developed by COLAUTTI [13] and REID [14], who obtained numerous inequalities of the extended Wirtinger type.

In the final part of the chapter, we shall establish some particular inequalities of this nature due to NORTHCOTT [27] and BELLMAN [28], and the more interesting discrete versions due to FAN, TAUSSKY, and TODD [29]. Related results are due to OSTROWSKI [20.1] and BLOCK [30].

As discussed in some detail by HARDY, LITTLEWOOD, and PÓLYA, and as we shall briefly sketch below, the problems we have been discussing can be treated by means of the calculus of variations; but for a number of reasons this is not a satisfactory procedure to employ, and little use will be made of it. Finally, we have omitted the results of BORG [31], extensions of the original result of LYAPUNOV, which are of great importance in the theory of linear differential equations with periodic coefficients; see STARZINSKII [32] for a survey of these matters, and many further results.

§ 2. Some Inequalities of B. Sz.-Nagy

Let us begin by discussing interesting inequalities of NAGY [1.16]. Analogous results were established by E. SCHMIDT [1] for the case of finite intervals.

Theorem 1. *Let $y(x)$ be a function defined over $[-\infty, \infty]$, for which the integrals*

$$J_a = \int_{-\infty}^{\infty} |y|^a \, dx\,, \quad K_p = \int_{-\infty}^{\infty} |y'|^p \, dx \tag{1}$$

exists for some $a > 0$ and some $p \geqq 1$. Then

$$\max_{-\infty < x < \infty} |y| \leqq \left(\frac{r}{2}\right)^{1/r} J_a^{(p-1)/(pr)} K_p^{1/(pr)}\,, \tag{2}$$

where $r = 1 + (p-1)\,a/p$; further, for $b > 0$,

$$J_{a+b} \leqq \left[\frac{q}{2} H\left(\frac{r}{b}\,\frac{p-1}{p}\right)\right]^{b/r} J_a^{1+b(p-1)/(pr)} K_p^{1/(br)}\,, \tag{3}$$

where

$$H(u, v) = \frac{(u+v)^{-(u+v)}\,\Gamma(1+u+v)}{u^{-u}\,\Gamma(1+u)\,\Gamma(1+v)}\,. \tag{4}$$

Proof. Consider first the case $p = 1$ and the inequality (2). Since J_a is finite, there exist sequences $\{a_n\}$, $\{b_n\}$, possessing the property that

$$y(a_n) \to 0\,, \quad y(b_n) \to 0 \quad \text{as} \quad a_n \to \infty\,, \quad b_n \to -\infty\,.$$

Then

$$\int_{-\infty}^{\infty} |y'|\,dx \geqq \mp \int_{z}^{\infty} y'\,dx \pm \int_{-\infty}^{z} y'\,dx = \lim_{n\to\infty}\left(\mp\int_{z}^{a_n} \pm \int_{a_n}^{z}\right) y'\,dx \geqq \pm 2y(z)\,. \tag{5}$$

Hence, for arbitrary z we have

$$\pm 2y(z) \leqq \int_{-\infty}^{\infty} |y'|\,dx\,, \tag{6}$$

the stated inequality for $p = 1$.

Now take $p > 1$. Using Hölder's inequality, we have

$$\begin{aligned}\int_{-\infty}^{\infty} |y|^{(p-1)\,a/p} |y'|\,dx &\leqq \left(\int_{-\infty}^{\infty} |y|^a\,dx\right)^{(p-1)/p} \left(\int_{-\infty}^{\infty} |y'|^p\,dx\right)^{1/p} \\ &= J_a^{(p-1)/p} K_p^{1/p}\,.\end{aligned} \tag{7}$$

Also, we have

$$\begin{aligned}\int_{-\infty}^{\infty} |y|^{(p-1)/p} |y'|\,dx &\geqq \left[-\int_0^{\infty} + \int_{-\infty}^{0}\right] (\operatorname{sgn} y)\,|y|^{(p-1)\,a/p} y'\,dx \\ &= \frac{1}{1+(p-1)\,a/p}\left[-\int_0^{\infty} + \int_{-\infty}^{0}\right] \frac{d}{dx}\left(|y|^{1+(p-1)\,a/p}\right) dx \\ &= \frac{1}{q}\left[2\,|y(0)|^q - \lim_{n\to\infty} |y(a_n)|^q - \lim_{n\to\infty} |y(b_n)|^q\right] = \frac{2}{q}\,|y(0)|^q\,.\end{aligned} \tag{8}$$

Since the integrals are invariant under the transformation $y(x) \to y(x+\lambda)$, we see that (2) holds.

The proof of (3) is more complicated and we refer the reader to the original paper [1.16].

§ 3. Inequalities Connecting u, u', and u''

Let us now derive some inequalities connecting u, u', and u''. We shall begin with some inclusion remarks and then show that these may be refined to yield inequalities.

Theorem 2. *If* $u \in L^p[0, \infty]$, $p \geqq 1$, $u'' \in L^r[0, \infty]$, $r \geqq 1$, *then* $u' \in L^m[0, \infty]$ *for* $m \geqq \max(p, r)$.

If $p = \infty$, *the condition* $u \in L^p[0, \infty]$ *is to be interpreted as* $\max_{t \geqq 0} |u| \leqq a < \infty$.

The two most interesting cases of this result, $p = r = m = \infty$, $p = r = m = 2$, were presented by BELLMAN in [1.3] of Chapter 4. We follow the method given there.

The bound $\max(p, r)$ can be improved. It was shown by NURENBERG, in unpublished work, that techniques similar to those used by NAGY, based on integration by parts and elementary estimates, can be used to obtain extensions of the following results.

Proof. Write

$$u'' - u = f + g\,, \tag{1}$$

where $f \in L^r$ and $g \in L^p$. Considering this relation to be a linear differential equation for u, with forcing term $f + g$, we can write u in the form

$$u = c_1 e^t + c_2 e^{-t} + \frac{1}{2}\int_0^t [e^{(t-s)} - e^{-(t-s)}]\,[f(s) + g(s)]\,ds\,. \tag{2}$$

The assumption that $f \in L^r$ and $g \in L^p$ implies that the integral

$$\int_0^\infty e^{-s}\,[f(s) + g(s)]\,ds$$

converges and that

$$e^{-t}\int_0^t e^s\,[f(s) + g(s)]\,ds$$

is bounded as $t \to \infty$. Hence in order that $u \in L^p$, it is necessary that

$$c_1 + \frac{1}{2}\int_0^\infty e^{-s}\,[f(s) + g(s)]\,ds = 0\,. \tag{3}$$

Using this relation, we find that (2) takes the form

$$u = c_2 e^{-t} - \frac{e^t}{2}\int_t^\infty e^{-s}\,[f + g]\,ds - \frac{e^{-t}}{2}\int_0^t e^s\,[f + g]\,ds\,, \tag{4}$$

which yields

$$u' = -c_2 e^{-t} - \frac{e^t}{2}\int_t^\infty e^{-s}\,[f + g]\,ds + \frac{e^{-t}}{2}\int_0^t e^s\,[f + g]\,ds\,. \tag{5}$$

From this representation, we readily derive the stated results. To begin with, consider the case in which $m = \infty$. Since, by HÖLDER's inequality,

$$
\begin{aligned}
\left|\int_t^\infty e^{-s} f\,ds\right| &\leqq \left(\int_t^\infty |f|^p\,ds\right)^{1/p} \left(\int_t^\infty e^{-qs}\,ds\right)^{1/q} \\
&\leqq \frac{e^{-t}}{q}\left(\int_0^\infty |f|^p\,ds\right)^{1/p}, \\
\left|\int_0^t e^{s} f\,ds\right| &\leqq \left(\int_0^t e^{qs}\,ds\right)^{1/q} \left(\int_0^t |f|^p\,ds\right)^{1/p} \\
&\leqq \frac{e^{t}}{q}\left(\int_0^\infty |f|^p\,ds\right)^{1/p},
\end{aligned}
\tag{6}
$$

with corresponding results for integrals involving g, we see that $|u|$ and $|u'|$ are uniformly bounded for $t \geqq 0$.

If $1 \leqq m < \infty$, we have

$$
\begin{aligned}
|u'(t)|^m \leqq c_4(m) \Bigg[e^{-mt} &+ e^{mt}\left(\int_t^\infty e^{-s}|f|\,ds\right)^m \\
&+ e^{mt}\left(\int_0^\infty e^{-s}|g|\,ds\right)^m \\
&+ e^{-mt}\left(\int_0^t e^{s}|f|\,ds\right)^m \\
&+ e^{-mt}\left(\int_0^t e^{s}|g|\,ds\right)^m \Bigg].
\end{aligned}
\tag{7}
$$

Consider a typical term on the right-hand side. We have

$$
\begin{aligned}
&\int_0^\infty e^{mt}\left(\int_t^\infty e^{-s}|f|\,ds\right)^{m/p} dt \\
&\leqq \int_0^\infty e^{mt}\left(\int_t^\infty e^{-qs/2}\,ds\right)^{m/q} \left(\int_0^\infty e^{-ps/2}|f|^p\,ds\right)^{m/q} dt \\
&\leqq c_5 \int_0^\infty e^{mt/2}\left(\int_0^\infty e^{-ps/2}|f|^p\,ds\right)^{m/p} dt\,.
\end{aligned}
\tag{8}
$$

Integrating by parts, we obtain

$$
\begin{aligned}
&\int_0^\infty e^{mt/2}\left(\int_0^\infty e^{-ps/2}|f|^p\,ds\right)^{m/p} dt \\
&= \frac{m}{p}\int_0^\infty e^{(m-p)t/2}\,|f|^p\left(\int_t^\infty e^{-ps/2}\,|f|^p\,ds\right)^{m/p-1} dt \\
&\leqq c_6\int_0^\infty |f|^p\left(\int_0^\infty |f|^p\,ds\right)^{m/p-1} dt \\
&= \frac{c_6 p}{m}\left(\int_0^\infty |f|^p\,ds\right)^{m/p} < \infty .
\end{aligned} \tag{9}
$$

We see that the condition $m \geqq p$ plays an essential role. This completes the proof.

Observe that we have actually established a stronger result than stated. It is not necessary that $u \in L^p$ and $u'' \in L^r$, but only that

$$
\begin{aligned}
u &= \sum_{k=1}^n f_k , \\
u'' &= \sum_{k=1}^n g_k ,
\end{aligned} \tag{10}
$$

with

$$f_k \in L^{p_k}, \quad g_k \in L^{r_k}, \quad p_k, r_k \geqq 1 ,$$

and

$$m \geqq \max_{1\leqq k\leqq n} [p_k, r_k] . \tag{11}$$

§ 4. Inequalities Connecting u, $u^{(k)}$, and $u^{(n)}$

To obtain a corresponding result for the more general triplet of derivatives, u, $u^{(k)}$, and $u^{(n)}$, $n > k > 1$, we use the equation

$$u^{(n)} - u = f + g \tag{1}$$

if $n = 4m + 2$ or n is odd, and

$$u^{(n)} + u = f + g \tag{2}$$

if $n = 4m$.

The reason for this change of equation lies in our desire to avoid the case in which the characteristic equation contains a root with zero real part. The result corresponding to Theorem 2 is the following:

Theorem 3. *If* $u \in L^p$, $u^{(n)} \in L^r$, $p, r \geqq 1$, $n > 1$, *then* $u^{(k)} \in L^m$ *for* $m \geqq \max [p, r]$ *and* $k = 0, 1, 2, \ldots, n-1$.

The proof follows the same lines as in § 3.

§ 5. Alternative Approach for u, u', and u''

If in place of the statement of Theorem 2 we wish to obtain actual inequalities, we can proceed as follows. From the relation

$$\frac{d}{dt}\left[e^{-t}(u'+u)\right] = e^{-t}(u''-u)\,, \tag{1}$$

for $u(t)$ satisfying the hypothesis of Theorem 2, we obtain the relation

$$u' = -u - e^{t}\int_t^\infty e^{-s}\,[u''-u]\,ds\,. \tag{2}$$

This result is also obtainable from (3.5) upon integrating by parts and observing that $c_2 = [u(0) - u'(0)]/2$.

Hence

$$\max_{t\geqq 0}|u'| \leqq 2\max_{t\geqq 0}|u| + \max_{t\geqq 0}|u''|\,. \tag{3}$$

Replacing $u(t)$ by $u(rt)$ for $r > 0$, we obtain the relation

$$r\max_{t\geqq 0}|u'| \leqq 2\max_{t\geqq 0}|u| + r^2\max_{t\geqq 0}|u''|\,, \tag{4}$$

for $r > 0$. From this it follows that

$$(\max_{t\geqq 0}|u'|)^2 \leqq 8\,(\max_{t\geqq 0}|u|)\,(\max_{t\geqq 0}|u''|)\,. \tag{5}$$

Similarly, we can obtain an inequality connecting $\int_0^\infty |u'|^m dt$, $\int_0^\infty |u|^p dt$, and $\int_0^\infty |u''|^r\,dt$; see the arguments given in [1.25]. To obtain a result corresponding to (5), we use the lemma of § 9, below. Unfortunately, this method does not yield best possible constants.

Perhaps the easiest way to obtain extensions of equation (2) is to use the vector-matrix relation

$$\frac{d}{dt}\,(e^{At}x) = e^{At}(Ax + x')\,, \tag{6}$$

where x is an n-dimensional vector, or

$$x = e^{-At}\int_t^\infty e^{As}(Ax + x')\,ds\,, \tag{7}$$

where A is a stability matrix.

Choosing A suitably, and

$$x = \begin{pmatrix} u \\ u' \\ \vdots \\ u^{(n-1)} \end{pmatrix}, \tag{8}$$

we obtain a variety of relations connecting $u^{(k)}$ with linear combinations of u and the derivatives of u. In this way, we can derive a number of extensions of (5).

§ 6. An Inequality of Halperin and von Neumann and Its Extensions

An extension of the foregoing techniques yields a generalization of the previous results.

Theorem 4. *If*

$$\begin{aligned} &\text{(a)} \quad u'' + a_1(t)\, u' + a_2(t)\, u \in L^p, \\ &\text{(b)} \quad u \in L^r, \\ &\text{(c)} \quad |a_1(t)|,\, |a_2(t)| \leqq c_1 < \infty\,, \quad 0 \leqq t\,, \end{aligned} \tag{1}$$

then $u, u' \in L^m$ *for* $m \geqq \max(r, p)$.

Proof. Let us discuss only the case in which $m = \infty$, which means that we are studying the uniform boundedness of $|u'|$. The general case can be treated in the same fashion. In addition to the device we have been using in the previous sections, we must introduce an additional one.

We write

$$\begin{aligned} u'' - M^2 u &= u'' + a_1(t)\, u' + a_2(t)\, u \\ &\quad - [a_1(t)\, u' + a_2(t)\, u + M^2 u] \\ &= f(t) - a_1(t)\, u', \end{aligned} \tag{2}$$

where $f(t) \in B\,[0, \infty]$, the space of functions uniformly bounded over $[0, \infty]$, by virtue of the assumptions in (1). Here M is a parameter to be chosen in an expeditious fashion.

Solving for u, we write

$$\begin{aligned} u = c_1 e^{-Mt} + c_2 e^{Mt} + \frac{1}{2M} \int_0^t \left[e^{M(t-s)} - e^{-M(t-s)}\right] f(s)\, ds \\ - \frac{1}{2M} \int_0^t \left[e^{M(t-s)} - e^{-M(t-s)}\right] a_1(s)\, u'(s)\, ds\,. \end{aligned} \tag{3}$$

Since it is easily established that the relations

$$[u'' + a_1(t)\, u' + a_2(t)\, u] \in B\,[0, \infty]\,, \quad a_i(t) \in B\,[0, \infty]\,, \quad i = 1, 2,$$

imply that

$$|u'(t)| \leqq c_3\, e^{bt}$$

for some constants c_3 and b, we see that the integrals

$$\int_0^\infty e^{-Ms}\, a_1(s)\, u'(s)\, ds \text{ and } \int_0^\infty e^{-Ms} f(s)\, ds$$

converge, provided that M is sufficiently large. Consequently (3) has the form

$$
\begin{aligned}
u = c_1 e^{-Mt} &- \frac{1}{2M} \int_0^t e^{-Mt} e^{Ms} f(s)\, ds \\
&+ \frac{1}{2M} \int_0^t e^{-Mt} e^{Ms} a_1(s)\, u'(s)\, ds \\
&+ \frac{e^{Mt}}{2M} \int_t^\infty e^{-Ms} [f(s) - a_1(s)\, u'(s)]\, ds\,,
\end{aligned} \tag{4}
$$

and accordingly

$$
\begin{aligned}
u' = -M c_1 e^{-Mt} &+ \frac{1}{2} \int_0^t e^{-Mt} e^{Ms} f(s)\, ds \\
&- \frac{1}{2} \int_0^t e^{-Mt} e^{Ms} a_1(s)\, u'(s)\, ds \\
&+ \frac{e^{Mt}}{2} \int_t^\infty e^{-Ms} [f(s) - a_1(s)\, u'(s)]\, ds\,.
\end{aligned} \tag{5}
$$

From this it follows for an appropriate constant c_2 that

$$
\begin{aligned}
|u'| \leqq c_2 \Bigg[e^{-Mt} &+ e^{-Mt} \int_0^t e^{Ms} |f(s)|\, ds \\
&+ e^{-Mt} \int_0^t e^{Ms} |u'(s)|\, ds \\
&+ e^{Mt} \int_t^\infty e^{-Ms} |f(s)|\, ds \\
&+ e^{Mt} \int_t^\infty e^{-Ms} |u'(s)|\, ds \Bigg]\,.
\end{aligned} \tag{6}
$$

Thus, for an appropriate c_3,

$$
\begin{aligned}
\max_{0 \leqq t \leqq T} |u'| &\leqq c_3 + \Big(\max_{0 \leqq t \leqq T} |u'| \Big) M^{-1} \\
&+ \max_{0 \leqq t \leqq T} \left(e^{Mt} \int_t^\infty e^{-Ms} |u'(s)|\, ds \right),
\end{aligned} \tag{7}
$$

whence, for $M > 1$,

$$
\begin{aligned}
\max_{0 \leqq t \leqq T} |u'| &\leqq c_3 (1 - 1/M)^{-1} \\
&+ \max_{0 \leqq t \leqq T} \left(e^{Mt} \int_0^\infty e^{-Ms} |u'(s)|\, ds \right) (1 - 1/M)^{-1}\,.
\end{aligned} \tag{8}
$$

Assuming now that $M \geqq 3$, we have

$$\max_{0 \leqq t \leqq T} |u'| \leqq c_4 + c_4 \max_{0 \leqq t \leqq T} \left(e^{Mt/2} \int_t^\infty e^{-Ms/2} |u'(s)| \, ds \right). \tag{9}$$

On the other hand, returning to (6), we obtain

$$\begin{aligned} \int_t^\infty e^{-Mt/2} |u'(t)| \, dt \leqq c_2 \Bigg[& \int_t^\infty e^{-3Mt/2} \, dt \\ & + \int_t^\infty e^{-3Mt/2} \left(\int_0^t e^{Ms} |f(s)| \, ds \right) dt \\ & + \int_t^\infty e^{-3Mt/2} \left(\int_0^t e^{Ms} |u'(s)| \, ds \right) dt \\ & + \int_t^\infty e^{Mt/2} \left(\int_t^\infty e^{-Ms} |f(s)| \, ds \right) dt \\ & + \int_t^\infty e^{Mt/2} \left(\int_t^\infty e^{-Ms} |u'(s)| \, ds \right) dt \Bigg]. \end{aligned} \tag{10}$$

Integrating by parts, we get

$$\begin{aligned} & \int_t^\infty e^{-Mt/2} |u'(t)| \, dt \\ & \leqq c_2 \Bigg[\frac{2}{3M} e^{-3Mt/2} + \frac{2}{3M} e^{-3Mt/2} \int_0^t e^{Ms} |f(s)| \, ds + \frac{2}{3M} \int_t^\infty e^{-Ms/2} |f(s)| \, ds \\ & + \frac{2}{3M} e^{-3Mt/2} \int_0^t e^{Ms} |u'(s)| \, ds + \frac{2}{3M} \int_t^\infty e^{-Ms/2} |u'(s)| \, ds \\ & + \frac{2e^{Mt/2}}{M} \int_t^\infty e^{-Ms} |f(s)| \, ds + \frac{2}{M} \int_t^\infty e^{-Mt/2} |f(t)| \, dt \\ & + \frac{2e^{Mt/2}}{M} \int_t^\infty e^{-Ms} |u'(s)| \, ds + \frac{2}{M} \int_t^\infty e^{-Ms/2} |u'(s)| \, ds \Bigg]. \end{aligned} \tag{11}$$

It follows that

$$e^{Mt/2} \int_t^\infty e^{-Mt/2} |u'(t)| \, dt \leqq c_3 \left[1 + \frac{e^{-Mt}}{M} \int_0^t e^{Ms} |u'(s)| \, ds \right]. \tag{12}$$

Hence

$$\max_{0 \leqq t \leqq T} \left[e^{Mt/2} \int_t^\infty e^{-Mt/2} |u'(t)| \, dt \right] \leqq c_3 + \frac{1}{M} \max_{0 \leqq t \leqq T} |u'(s)|. \tag{13}$$

Combining (8) and (13), we see that $\max_{0 \leq t \leq T} |u'(s)|$ is uniformly bounded for $T \geqq 0$. This completes the proof.

The proof of the general result for linear differential operators of arbitrary order proceeds in similar fashion.

§ 7. Results Analogous to Those of Sz.-Nagy

Let us indicate briefly how we can obtain results similar to those of NAGY given in § 2.

Write

$$u' + u = f + g\,, \tag{1}$$

where $f \in L^p[-\infty, \infty]$, $g \in L^r[-\infty, \infty]$, $p, r \geqq 1$. Then, arguing as above, we find that u must have the form

$$u = e^{-t} \int_{-\infty}^{t} f(s)\, e^s\, ds + e^{-t} \int_{-\infty}^{t} g(s)\, e^s\, ds\,. \tag{2}$$

The conditions imposed upon f and g yield the uniform boundedness of u. Hence $u \in L^p(-\infty, \infty)$ and $u' \in L^r(-\infty, \infty)$ imply that $u \in L^{p+b}[-\infty, \infty]$ for $b \geqq 0$. This result, however, is not as strong as that given in § 2.

§ 8. Carlson's Inequality

An inequality due to CARLSON [1.19] is the following.

Theorem 5. *If $g(t) \geqq 0$ and the integrals on the right exist, then*

$$\int_0^\infty g(t)\, dt \leqq \sqrt{\pi} \left\{\int_0^\infty [g(t)]^2\, dt\right\}^{1/4} \left\{\int_0^\infty [t g(t)]^2\, dt\right\}^{1/4}. \tag{1}$$

As pointed out by NAGY [1.16], the result follows from Theorem 1 when we take

$$f(x) = \sqrt{\frac{2}{\pi}} \int_0^\infty g(t) \cos xt\, dt\,, \tag{2}$$

$p = a = 2$, and apply the Parseval-Plancherel formula [1].

§ 9. Generalizations of Carlson's Inequality

CARLSON's inequality can be substantially extended. The general problem is that of determining bounds for $\int_0^\infty f(x)\, dx$, given the existence of $\int_0^\infty x^a f^b\, dx$ and $\int_0^\infty x^c f^d\, dx$. Results in this direction have been obtained by HARDY [1.21], GABRIEL [1.22], BEURLING [1.23], CATON [1.24], SZ. NAGY [1.16], BELLMAN [1.25], and KJELLBERG [1.20]. As indicated by KJELLBERG, this is a moment problem and can thus be treated by the techniques described in Chapter 3. See also BOAS [1].

Here we shall present, following [1.25], a simple technique for obtaining results of this type when optimum bounds are not of interest. For a derivation of the best inequality, see LEVIN [1.26]. Let us demonstrate the following result:

Theorem 6. *If* $f(x) \geqq 0$ *and* $p, q > 0$, $0 < \lambda < p + 1$, $0 < \mu < q + 1$, *then*

$$\left(\int_0^\infty f(x)\,dx\right)^{p\mu + q\lambda + \mu + \lambda} \leqq K(p, q, \lambda, \mu)\left(\int_0^\infty x^{p-\lambda} f^{p+1}\,dx\right)^{\mu}\left(\int_0^\infty x^{q+\mu} f^{q+1}\,dx\right)^{\lambda}. \tag{1}$$

Proof. We write

$$\int_0^\infty f(x)\,dx = \int_0^\infty \frac{x^{(p-\lambda)/(p+1)} f\,dx}{x^{(p-\lambda)/(p+1)}(1+x)} + \int_0^\infty \frac{x^{(q+\mu)/(q+1)} f\,dx}{x^{(q+\mu)/(q+1)}(1+1/x)}\,. \tag{2}$$

Then, using HÖLDER's inequality, we have

$$\begin{aligned}
\int_0^\infty \frac{x^{(p-\lambda)/(p+1)} f\,dx}{x^{(p-\lambda)/(p+1)}(1+x)} &\leqq \left(\int_0^\infty x^{p-\lambda} f^{p+1}\,dx\right)^{1/(p+1)} \\
&\quad\cdot\left(\int_0^\infty \frac{dx}{x^{(1-\lambda)/p}(1+x)^{1+1/p}}\right)^{1/(p+1)}, \\
\int_0^\infty \frac{x^{(q+\mu)/(q+1)} f\,dx}{x^{(q+\mu)/(q+1)}(x+1/x)} &\leqq \left(\int_0^\infty x^{q+\mu} f^{q+1}\,dx\right)^{1/(q+1)} \\
&\quad\cdot\left(\int_0^\infty \frac{dx}{x^{(1+\mu)/q}(1+1/x)^{1+1/q}}\right)^{1/(q+1)}.
\end{aligned} \tag{3}$$

We thus obtain the preliminary result that

$$\begin{aligned}
\int_0^\infty f(x)\,dx &\leqq c_1\left(\int_0^\infty x^{p-\lambda} f^{p+1}\,dx\right)^{1/(p+1)} \\
&\quad + c_2\left(\int_0^\infty x^{q+\mu-1} f^{q+1}\,dx\right)^{1/(q+1)},
\end{aligned} \tag{4}$$

where c_1 and c_2 are the integrals independent of f in (3).

Let us now employ a device that is useful in many similar situations. When $f(x)$ is replaced by $f(x/t)$, $t > 0$, (4) yields, after a change of variable,

$$\begin{aligned} t\int_0^\infty f(x)\,dx \leqq c_1 t^{(p-\lambda+1)/(p+1)} \left(\int_0^\infty x^{p-\lambda} f^{p+1}\,dx\right)^{1/(p+1)} \\ + c_2 t^{(q+\mu+2)/(q+1)} \left(\int_0^\infty x^{q+\mu+1} f^{q+1}\,dx\right)^{1/(q+1)}, \end{aligned} \tag{5}$$

for all $t > 0$.

To obtain the stated result, we employ the following lemma:

Lemma. *If $v > u > 0$, $a, b, c > 0$, and*

$$b x^u \leqq c + a x^v \tag{6}$$

for all $x > 0$, then

$$c^{v-u} a^u \geqq k(u, v)\, b^v. \tag{7}$$

The lemma follows upon setting $x = (bu/av)^{1/(v-u)}$, and this completes the proof of the theorem. The result is an extension of the well-known fact that $b^2 \leqq 4ac$ if $bx \leqq c + ax^2$ for all $x \geqq 0$, and $a, b, c \geqq 0$.

It is clear that corresponding inequalities may be obtained for infinite series by means of the same techniques.

§ 10. Wirtinger's Inequality and Related Results

In a number of cases, in addition to obtaining inclusion results and explicit inequalities, we are able to derive precise inequalities, together with a determination of the class of functions for which equality occurs. In general, this involves a certain amount of ingenuity, and occasionally an appeal to the apparatus of the calculus of variations. Sometimes, however, the results can readily be obtained in a number of ways. This is particularly true of inequalities involving quadratic functionals.

Perhaps the most interesting example, and one that is widely used in analysis, is the Wirtinger inequality; see Hardy, Littlewood, and Pólya ([1.1] of Chapter 1) Blaschke [1], and Fan, Taussky, and Todd [1.29].

Theorem 7. *If $u(t)$ has period 2π, and*

$$\int_0^{2\pi} u(t)\,dt = 0 \tag{1}$$

then

$$\int_0^{2\pi} [u(t)]^2\,dt \leqq \int_0^{2\pi} [u'(t)]^2\,dt, \tag{2}$$

with strict inequality unless

$$u(t) = c_1 \cos t + c_2 \sin t. \tag{3}$$

Let us give three proofs, the first based on the theory of Fourier series, a method used by HURWITZ [2] to establish the isoperimetric inequality in the plane and in space, the second on Sturm-Liouville theory, and the third on an integral inequality. This last method was extended by BEESACK [1.12]. The three proofs are given in §§ 11—13, below.

We shall discuss the isoperimetric inequality in detail in our second volume. For extensions of this inequality, see [1.16] and the paper by SCHOENBERG ([48.7] of Chapter 2) to which we have already referred in connection with determinantal inequalities.

For a discussion of the related inequality of WEYL, see SLEPIAN and POLLAK [3], LANDAU and POLLAK [4], and WHITTLE [5].

§ 11. Proof Using Fourier Series

Let us write, for $u(t)$ a real function,

$$u(t) \sim \sum_{n \neq 0} a_n e^{int}, \tag{1}$$

where the term involving $n = 0$ does not appear, by virtue of the condition (10.1). Then

$$u'(t) \sim \sum_{n \neq 0} i n a_n e^{int}, \tag{2}$$

and PARSEVAL's relation yields

$$\begin{aligned} \int_0^{2\pi} [u(t)]^2\, dt &= \sum_{n \neq 0} |a_n|^2, \\ \int_0^{2\pi} [u'(t)]^2\, dt &= \sum_{n \neq 0} n^2 |a_n|^2. \end{aligned} \tag{3}$$

From these, we obtain the desired relationship (10.2), with strict inequality unless

$$a_n = 0, \quad |n| \geqq 2. \tag{4}$$

§ 12. Sturm-Liouville Theory

Let $p(t)$ be a bounded positive function, and consider the problem of obtaining an inequality connecting

$$\int_0^{2\pi} p[u(t)]^2\, dt \quad \text{and} \quad \int_0^{2\pi} [u'(t)]^2\, dt.$$

Let us assume that $u(0) = u(2\pi) = 0$, and ask for the minimum of $\int_0^{2\pi} [u'(t)]^2\, dt$, subject to the normalization

$$\int_0^{2\pi} p u^2\, dt = 1. \tag{1}$$

Then, proceeding formally, and using a Lagrange multiplier, we are led to the problem of minimizing the quadratic functional

$$J(u) = \int_0^{2\pi} (u'^2 - \lambda p u^2)\, dt, \tag{2}$$

over all $u(t)$ for which the integral exists, and which satisfy (1) and the foregoing end-point conditions.

The Euler equation is

$$u'' + \lambda p u = 0. \tag{3}$$

If u satisfies this equation, we have

$$0 = \int_0^{2\pi} u\,(u'' + \lambda p u)\, dt = [uu']_0^{2\pi} - \int_0^{2\pi} [u'(t)]^2\, dt + \lambda \int_0^{2\pi} p u^2\, dt, \tag{4}$$

or

$$\lambda = \int_0^{2\pi} u'(t)^2\, dt. \tag{5}$$

Hence, the required minimum value is the smallest characteristic value of the Sturm-Liouville problem associated with (1) and the two-point boundary values

$$u(0) = u(2\pi) = 0. \tag{6}$$

For a discussion of problems of this nature, see Ince [1] and Coddington-Levinson ([34.3] of Chapter 4).

§ 13. Integral Identities

A third approach to these problems, and to a variety of similar problems involving higher-order derivatives, is based on a class of identities due to Beesack [1.12]; see also Hardy, Littlewood, and Pólya ([1.1] of Chapter 1).

Let $v(t)$ be a real function satisfying the Riccati equation

$$v' + v^2 + p(t) = 0, \tag{1}$$

with an initial condition we shall subsequently discuss. Then, proceeding formally, we get

$$\begin{aligned}\int_0^{2\pi} [u' - v\,(u-b)]^2\, dt &= \int_0^{2\pi} u'^2\, dt + \int_0^{2\pi} v^2 (u-b)^2\, dt \\ &\quad - 2\int_0^{2\pi} v u' (u-b)\, dt.\end{aligned} \tag{2}$$

Integrating by parts, we see that

$$\begin{aligned}-\int_0^{2\pi} 2 v u' (u-b)\, dt &= [-v\,(u-b)^2]_0^{2\pi} + \int_0^{2\pi} v' (u-b)^2\, dt \\ &= [-v\,(u-b)^2]_0^{2\pi} - \int_0^{2\pi} (v^2 + p)\,(u-b)^2\, dt.\end{aligned} \tag{3}$$

Using this result in (2), we obtain the identity

$$\int_0^{2\pi} [u' - v\,(u-b)]^2\,dt = \int_0^{2\pi} [u'^2 - p\,(u-b)^2]\,dt + [-v\,(u-b)^2]_0^{2\pi}. \tag{4}$$

Suppose that $v\,(u-b)^2$ is zero at $t = 0$, and that $v\,(t)$ exists for $0 \leqq t \leqq 2\pi$ and is nonnegative there. Then (4) yields

$$\int_0^{2\pi} [u' - v\,(u-b)^2]\,dt + v\,(2\pi)\,[u\,(2\pi) - b]^2 = \int_0^{2\pi} [u'^2 - p\,(u-b)^2]\,dt\,. \tag{5}$$

A special case of this identity is given in HARDY, LITTLEWOOD, and PÓLYA to establish WIRTINGER's inequality. In his paper, BEESACK [1.11] gives an analogue of this result for fourth-order equations of the form

$$u^{(iv)} \pm p\,u = 0\,, \tag{6}$$

and uses this to establish a number of inequalities connecting $\int_0^{2\pi} u''^2\,dt$ and $\int_0^{2\pi} p\,(t)\,u^2\,dt$. He also discusses the rigorous derivation of (4). Further results for ordinary and partial differential equations are given in REID [1.14] with application to isoperimetric problems.

§ 14. Colautti's Results

As might be expected, the identity (13.6) associated with the self-adjoint equation $u'' + p\,u = 0$ is a special case of the identities that can be obtained with the aid of a linear differential equation $L\,(u)$ and its adjoint equation $L^*(u) = 0$. This theme has been extensively and adroitly developed by COLAUTTI [1.13], who used the resultant identities to derive a variety of interesting inequalities of the type we have been discussing. Related results are given in HORMANDER [1].

§ 15. Partial Differential Equations

BEESACK [1.11] pointed out in passing that identities similar to (13.5) can be derived for partial differential equations. Thus, if

$$v_{xx} + v_{yy} + p\,v = 0 \tag{1}$$

for $(x, y) \in R$, and $v = 0$ on the boundary B of R, then we have

$$\int_R (u_x^2 + u_y^2 - p\,u^2)\,dA = \int_R [(u_x - w\,u)^2 + (u_y - z\,u)^2]\,dA\,, \tag{2}$$

where

$$w = \frac{v_x}{v}, \quad z = \frac{v_y}{v}, \tag{3}$$

provided that $u = 0$ on R, and that the various functions and integrals exist. Similarly, the techniques of Colautti [1.13] can be extended to partial differential equations. Note that the "Riccati equation" is now

$$w_x + z_y + w^2 + z^2 + p = 0. \tag{4}$$

§ 16. Matrix Version

A systematic way of obtaining analogues of (13.5) for higher-order linear differential equations, and for linear operator equations in general, is to use vector-matrix notation. Let x and b be n-dimensional vectors, x variable and b constant, and Z an n by n symmetric matrix satisfying the equation

$$Z' + Z^2 + P(t) = 0. \tag{1}$$

Then, proceeding as above, we get

$$\begin{aligned} &\int_0^{2\pi} (x' - Z(x-b),\ x' - Z(x-b))\,dt \\ &= \int_0^{2\pi} (x', x')\,dt + \int_0^{2\pi} (Z(x-b),\ Z(x-b))\,dt \\ &- 2\int_0^{2\pi} (x', Z(x-b))\,dt. \end{aligned} \tag{2}$$

Since

$$\frac{d}{dt}((x-b), Z(x-b)) = 2(x', Z(x-b)) + (x-b, Z'(x-b)), \tag{3}$$

we have, after the same manipulations as in the scalar case, the identity

$$\begin{aligned} \int_0^{2\pi} [(x', x') - (x, P(t)\,x)]\,dt &= \int_0^{2\pi} (x' - Z(x-b),\ x' - Z(x-b))\,dt \\ &+ (x-b, Z(x-b))_{t=2\pi}, \end{aligned} \tag{4}$$

provided that $(x-b, Z(x-b))$ vanishes at $t = 0$, and $Z(2\pi)$ is positive definite.

To use this identity to obtain analogues of (13.5) for higher-order linear differential equations, we first convert the equation into a second-order vector-matrix equation,

$$x'' + P(t)\,x = 0, \tag{5}$$

use (4), and then convert back. Generally, we shall have to use both the original linear differential equation and its adjoint.

§ 17. Higher Derivatives and Higher Powers

Having obtained precise results for squares and first derivatives, we naturally turn to the problem of relating the two means

$$\int_0^{2\pi} [u(t)]^{2r}\, dt, \int_0^{2\pi} [u^{(k)}(t)]^{2r}\, dt\,. \tag{1}$$

Problems of this type, although simple in principle as far as variational analysis is concerned, pose certain difficuties if precise inequalities are desired. A detailed analysis of many problems of this type may be found in the previously cited book by HARDY, LITTLEWOOD, and PÓLYA, and in papers by KOLMOGOROFF [1.10] and BEESACK [1.11], [1,12].

Here we wish to present two readily derived results due respectively to NORTHCOTT [1.27] and BELLMAN [1.28]; see also OSTROWSKI [20.1].

Theorem 8. *If* $u(t) = u(t + 2\pi)$, *if* $u(t)$, $u'(t)$, $\ldots$, $u^{(k-1)}(t)$ *are absolutely continuous, and if*

$$\int_0^{2\pi} u(s)\, ds = 0\,,$$

then

$$\max_{0 \leqq t \leqq 2\pi} |u(t)| \leqq a_k \max_{0 \leqq t \leqq 2\pi} |u^{(k)}(t)|\,, \tag{2}$$

where $\{a_k\}$ *is the following prescribed sequence:*

$$a_1 = \frac{\pi}{2}\,, \quad a_2 = \frac{\pi^2}{8}\,, \quad a_3 = \frac{\pi^3}{24}\,, \quad \ldots \tag{3}$$

This result is the best possible in the sense that for any k, *there is a function* $u(t)$ *for which equality is attained.*

Theorem 8 may be considered a limiting case of the following result:

Theorem 9. *If* $u(t) = u(t + 2\pi)$, *if* $u(t)$, $u'(t)$, $\ldots$, $u^{(k-1)}(t)$ *all exist, if* $u^{(k-1)}(t)$ *is the integral of a function in* $L^{2r}[0, 2\pi]$, *and if*

$$\int_0^{2\pi} u(s)\, ds = 0\,,$$

then

$$\int_0^{2\pi} [u(t)]^{2r}\, dt \leqq a_k^{2r} \int_0^{2\pi} [u^{(k)}(t)]^{2r}\, dt\,. \tag{4}$$

Here the results are no longer the best possible.

§ 18. Discrete Versions of Fan, Taussky, and Todd

As is frequently the case, discrete versions of inequalities are more interesting in that they are usually more difficult to establish and require more effort in the determination of best constants and extremal functions.

An early result in this direction is due to BLASCHKE [10.1]:

Theorem 10. *If L is the length of an equilateral polygon with n sides and area A, then*

$$L^2 \geqq \left[4n \tan \frac{\pi}{n}\right] A\,, \tag{1}$$

with strict inequality unless the polygon is regular.

Let us now discuss a number of results due to FAN, TAUSSKY, and TODD [1.29]. These results could equally well have been part of Chapter 2, which was devoted to inequalities pertaining to matrices. Their proofs depend on a detailed analysis of the characteristic roots and vectors of various special classes of symmetric matrices.

Theorem 11. *If $x_1, x_2, \ldots, x_n$ are n real numbers, $x_1 = 0$, then*

$$\sum_{i=1}^{n-1} (x_i - x_{i+1})^2 \geqq 4 \sin^2 \frac{\pi}{2\,(2n-1)} \sum_{i=2}^{n} x_i^2\,, \tag{2}$$

with equality only if $x_i = c\,x_i'$, $i = 1, 2, \ldots, n$, where

$$x_i' = \sin \frac{(i-1)\,\pi}{2n-1}\,, \quad i = 1, 2, \ldots, n\,. \tag{3}$$

The structure of the symmetric matrices arising in this and related variational problems discussed by FAN, TAUSSKY, and TODD had been investigated by RUTHERFORD [1] because of their great importance in a number of mathematical models of chemical and physical processes.

In addition to the foregoing result, a number of related results are given, corresponding to different conditions imposed upon x_1 and x_n.

§ 19. Discrete Case — Second Differences

A further analysis of the same kind yields the following result; see the paper by FAN, TAUSSKY, and TODD [1.29] referred to above.

Theorem 12. *If $x_1, x_2, \ldots, x_n$ are n real numbers, where $x_0 = x_{n+1} = 0$, then*

$$\sum_{i=0}^{n-1} (x_i - 2\,x_{i+1} + x_{i+2})^2 \geqq 16 \sin^4 \frac{\pi}{2\,(n+1)} \sum_{i=0}^{n} x_i^2\,, \tag{1}$$

with equality if and only if

$$x_i = c \sin \frac{\pi i}{n+1}\,, \quad i = 1, 2, \ldots, n\,. \tag{2}$$

There are analogous results given for other boundary conditions.

§ 20. Discrete Versions of Northcott-Bellman Inequalities

Let us now consider the discrete versions of some of the results given in § 17. The first result is an analogue of Theorem 9:

Theorem 13. *If $x_1, x_2, \ldots, x_n$ are n complex numbers such that*

$$\sum_{i=1}^{n} x_i = 0\,, \tag{1}$$

then, for any positive integer r, we have

$$\sum_{i=1}^{n} |x_i|^r \leqq [a_1(n)]^r \sum_{i=1}^{n} |x_i - x_{i+1}|^r \tag{2}$$

and

$$\sum_{i=1}^{n} |x_i|^r \leqq [a_2(n)]^r \sum_{i=1}^{n} |x_{i-1} - 2x_i + x_{i+1}|^r\,, \tag{3}$$

where

$$x_0 = x_n, \quad x_1 = x_{n+1}, \tag{4}$$

and

$$a_1(n) = \frac{n-1}{2}\,, \quad a_2(n) = \frac{n^2-1}{12}\,. \tag{5}$$

The inequalities (2) *and* (3) *are the best possible of this form.*

The second result is an analogue of Theorem 8, and of a continuous result due to Ostrowski [1]:

Theorem 14. *If n $(n \geqq 2)$ real numbers vary under the conditions*

$$\begin{aligned} &\text{(a)} \quad \sum_{i=1}^{n} x_i = 0\,, \\ &\text{(b)} \quad \max_{1 \leqq i \leqq n} |x_i| = 1\,, \\ &\text{(c)} \quad x_{n+1} = x_1\,, \end{aligned} \tag{6}$$

then the minimum of

$$J(x) = \max_{1 \leqq i \leqq n} |x_i - x_{i+1}| \tag{7}$$

is

$$\begin{aligned} &\frac{4}{n} \textit{ if } n \textit{ is even}, \\ &\frac{4n}{n^2-1} \textit{ if } n \textit{ is odd}. \end{aligned} \tag{8}$$

Proofs of these results, as well as some generalizations, may be found in the paper by Fan, Taussky, and Todd [1.29].

§ 21. Discussion

Results of the type presented in §§ 18—20 are important in connection with the numerical integration of ordinary and partial differential equations. Thus, for example, if a principal characteristic value for a region R is defined as the minimum of the quotient

$$J(u) = \frac{\iint\limits_R [u_x^2 + u_y^2]\,dx\,dy}{\iint\limits_R u^2 dx\,dy}, \tag{1}$$

over functions u satisfying the condition

$$u = 0 \text{ on } B, \text{ the boundary of } R, \tag{2}$$

(see [44.1] of Chapter 1 for a systematic discussion of problems of this nature), it is essential to know the connection between $\min\limits_u J(u)$ and the corresponding constant obtained from a discrete version of the problem.

Questions of this nature have also been discussed by FORSYTHE [1], KREIN [2], and SCHWARZ [3].

Bibliographical Notes

§ 1. A systematic theory of the growth of derivatives of functions is contained in the theory of quasi-analytic functions; see

1. MANDELBROJT, S.: Séries de Fourier et classes quasi-analytiques de fonctions. Paris: Gauthier-Villars 1935.

Many further references can be found in the book

2. BOAS, R. P.: Entire functions. New York: Academic Press, Inc. 1954.
3. SCHAEFFER, A. C.: Inequalities of A. MARKOFF and S. BERNSTEIN for polynomials and related functions. Bull. Am. Math. Soc. **47**, 565—579 (1941).
4. SZ.-NAGY, B.: Über gewisse Extremalfragen bei transformierten trigonometrischen Entwicklungen. I.: Periodischer Fall. II.: Nichtperiodischer Fall. Ber. Math. Phys. Kl. Sächs. Akad. Wiss. Leipzig **90**, 103—134 (1938); **91**, 3—24 (1939).

For results concerning partial differential operators, see [26.1] of Chapter 2 and the recent paper,

5. ARONSZAJN, N.: A unique continuation theorem for solutions of elliptic partial differential equations or inequalities of second order. J. Math. Pure Appl. (9) **36**, 235—249 (1957).
6. NIRENBERG, L.: On elliptic partial differential equations. Ann. Scuola Norm. Sup. Pisa (3) **13**, 115—162 (1959).
7. FRIEDRICHS, K. O.: Symmetric positive linear differential equations. Comm. Pure Appl. Math. **9**, 333—418 (1958).

For a discussion of the results of HADAMARD, references and extensions, see [1.1] of Chapter 1 and

8. TITCHMARSH, E. C.: Theory of functions. London: Oxford University Press 1939.
9. LANDAU, E.: Einige Ungleichungen für zweimal differentierbare Funktionen. Proc. London Math. Soc. **13**, 43—49 (1913).
10. KOLMOGOROFF, A.: On inequalities between upper bounds of the successive derivatives of an arbitrary function on an infinite interval. Am. Math. Soc. Translation No. 4. New York: American Mathematical Society 1949.
11. BEESACK, P. R.: Isoperimetric inequalities for the nonhomogeneous clamped rod and plate. J. Math. and Mech. **8**, 471—482 (1959).
12. BEESACK, P. R.: Integral inequalities of the Wirtinger type. Duke Math. J. **25**, 477—498 (1948).

References to work by PUCCI and FICHERA will be found in

13. COLAUTTI, M. P.: Sulla maggiorazione "a priori" delle soluzione delle equazione e del sistemi di equazioni differenziali lineari ordinarie del secondo ordine. Matematiche, Catania **11**, 8—99 (1956).
14. REID, W. T.: The isoperimetric inequality and associated boundary problems. J. Math. and Mech. **8**, 897—906 (1959).
15. BELLMAN, R.: On inequalities for differential operators. Proc. Am. Math. Soc. **9**, 589—597 (1958).
16. SZ.-NAGY, B.: Über Integralgleichungen zwischen einer Funktion und ihrer Ableitungen. Acta Sci. Math. Szeged **10**, 64—74 (1941).

The HALPERIN-VON NEUMANN result may be found in

17. HALPERIN, J.: Closures and adjoints of linear differential operators. Ann. of Math. **38**, 880—919 (1937).

Extensions are in

18. HALPERIN, J., and H. R. PITT: Integral inequalities associated with differential operators. Duke Math. J. **4**, 613—625 (1938).

CARLSON's inequality has attracted a considerable amount of attention; see

19. CARLSON, F.: Une inégalité. Arkiv Mat., Astr. och Fysik **25 B**, No. 1, 1—5 (1934).
20. KJELLBERG, B.: On some inequalities. C. R. Dixième Congrès des Mathématiciens Scandinaves 1946, pp. 333—340. Copenhagen: Jul. Gjellerups Forlag 1946.
21. HARDY, G. H.: A note on two inequalities. J. London Math. Soc. **11**, 167—170 (1936).
22. GABRIEL, R. M.: An extention of an inequality due to CARLSON. J. London Math. Soc. **12**, 130—132 (1937).
23. BEURLING, A.: Sur les intégrales de FOURIER absolument convergentes et leur application à une transformation fonctionnelle. Neuvième Congrès des Mathématiciens Scandinaves, 1938, pp. 345—366. Helsingfors 1939.
24. CATON, W. B.: A class of inequalities. Duke Math. J. **6**, 442—461 (1940).
25. BELLMAN, R.: An integral inequality. Duke Math. J. **9**, 547—550 (1943).

The problem of determining the best constant in CARLSON's inequality has been resolved by LEVIN. See the appendix to the Russian version of HARDY, LITTLEWOOD, and PÓLYA ([1. 1.] of Chapter 1), and

26. LEVIN, V. I.: Exact constants in inequalities of the Carlson type (Russian). Doklady Akad. Nauk SSSR (N. S.) **59**, 635—638 (1948).
27. NORTHCOTT, D. G.: Some inequalities between periodic functions and their derivatives. J. London Math. Soc. **14**, 198—202 (1939).
28. BELLMAN, R.: A note on periodic functions and their derivatives. J. London Math. Soc. **18**, 140—142 (1943).

For a detailed analysis of discrete versions, see

29. FAN, K., O. TAUSSKY and J. TODD: Discrete analogues of inequalitics of WIRTINGER. Monatsh. Math. Physik **59**, 73—90 (1955).

References to a number of other results may be found in [29], in particular, to work of OSTROWSKI, BÜCKNER, and PÓLYA.

30. BLOCK, H. D.: A class of inequalities. Proc. Am. Math. Soc. **8**, 844—851 (1957).

For an excellent exposition of the theory of linear differential equations with periodic coefficients and the connections between this field and inequalities involving differential operators, see

31. BORG, G.: Über die Stabilität gewisser Klassen von linearen Differentialgleichungen. Ark. Math., Astr. och Fysik **31 A**, 460—482 (1944).

32. STARZINSKII, V. M.: Survey of works on conditions of stability of the trivial solution of a system of linear differential equations with periodic coefficients. Translation Am. Math. Soc. (2) 1, 189—239. New York: American Mathematical Society 1955.

For inequalities relating integrals of trigonometric polynomials, see the papers of Sz.-NAGY in the bibliography of [1.2].

The idea of connecting stability theory with inequalities of the type considered by BORG stems from BEURLING.

§ 2.

1. SCHMIDT, E.: Über die Ungleichung, welche die Integrale über eine Potenz einer Funktion und über eine andere Potenz ihrer Ableitung verbindet. Math. Ann. 117, 301—326 (1940).

§ 3. See [1.3] of Chapter 4.

§ 6. The method is that given in [1.15].

§ 8. See

1. TITCHMARSH, E. C.: Introduction to the theory of Fourier integrals, Chapter 3. London: Oxford University Press 1937.

§ 9. The proof given here follows [1.25]. A number of related results are given in

1. BOAS, R. P.: Inequalities for monotonic series. J. Math. Analys. and Appl., to appear,

generalizing theorems of LOJASIEWICZ and KONYUSHKOV. Typical is the following result: If $a_n \geqq 0$, $\lambda \geqq 0$, $n^{-\lambda} a_n \downarrow$, $b > 1$, then

$$\left(\sum_{k=M}^{N} a_k\right)^p \geqq A_1(p, \lambda, b) \sum_{k=Mb}^{N} a_k^p k^{p-1}, p > 1,$$

$$\left(\sum_{k=M}^{N} a_k\right)^p \leqq A_2(p, \lambda, b) \sum_{k=M/b}^{N} a_k^p k^{p-1}, 0 < p < 1.$$

Note the analogy to the result by Gauss quoted at the end of this bibliography.

§ 10. See

1. BLASCHKE, W.: Kreis und Kugel, pp. 13—20. Leipzig: Veit u. Co. 1916; New York: Chelsea Publishing Co. 1949,

for some interesting geometrical applications of discrete versions. See also [48.7] of Chapter 2 and

2. HURWITZ, A.: Sur le problème des isoperimetres. C. R. Acad. Sci. (Paris) 132, 401—403 (1901) (Werke, I, 490—491).

The inequality of WEYL (Theorem 226, p. 165 of [1.1] in Chapter 1),

$$\left(\int_0^\infty f^2\,dx\right)^2 \leqq 4\left(\int_0^\infty x^2 f^2\,dx\right)\left(\int_0^\infty f'^2\,dx\right),$$

is a quantitiative expression of the HEISENBERG uncertainty principle of quantum mechanics. For further "uncertainty principles," see

3. SLEPIAN, D., and H. POLLAK: Prolate spheroidal wave functions, Fourier analysis and uncertainty I. Bell System Tech. J. 40, 43—64 (1961),

4. LANDAU, H. J., and H. POLLAK: Prolate sheroidal wave functions, Fourier analysis and uncertainty II. Bell System Tech. J. 40, 65—84 (1961).

See also

5. WHITTLE, P.: Continuous generalizations of TCHEBYCHEFs inequality. Probability Theory and Appl. (Russian) 3, 385—394 (1960).

§ 12.

1. INCE, E. L.: Ordinary differential equations. New York: Dover Publications 1944.

§ 14.

1. HORMANDER, L.: On the regularity of solutions of boundary problems. Acta Math. **99**, 225—264 (1958).

§ 18.

1. RUTHERFORD, D. R.: Some continuant determinants arising in physics and chemistry, I, II. Proc. Roy. Soc. Edinburgh, **62 A**, 229—236 (1947); **63** A, 232—241 (1952).

§ 20.

1. OSTROWSKI, A. M.: Über die Absolutabweichung einer differentierbaren Funktion von ihrem Integralmittelwert. Comment. Math. Helv. **10**, 226—227 (1937).

§ 21.

See [44.1] of Chapter 1 and

1. FORSYTHE, G.: Asymptotic lower bounds for the frequencies of certain polygonal membranes. Pacific J. Math. **4**, 467—480 (1954).
2. KREIN, M. G.: On certain problems in the maximum and minimum of characteristic values and on the Lyapunov zones of stability. Am. Math. Soc. Translations (2) **1**, 163—187. New York: American Mathematical Society 1955.
3. SCHWARZ, B.: On the extrema of the frequencies of nonhomogeneous strings with equimeasurable density. AFOSR TN 60—281. Haifa, Israel: Israel Institute of Technology 1959.

As we have indicated many times before, we are omitting a large number of interesting inequalities arising from probability theory and statistics. One of historical interest is due to GAUSS:

$$y^2 \int_y^\infty u(x)\, dx \leqq \frac{4}{9} \int_0^\infty x^2 u(x)\, dx ,$$

if $y > 0$ and $u(x)$ is nonincreasing. See

4. CRAMÉR, H.: Methods of mathematical statistics, p. 256, ex. 4. Princeton, N. J.: Princeton University Press 1946.

The functional-equation technique of dynamic programming can be used to obtain a nonlinear partial differential equation for the function $f(u, v, t)$ defined by

$$f(u, v, t) = \max_g \int_0^t x^{a_1} g^{b_1}\, dx ,$$

where $g(x)$ is subject to the constraints

$$\int_0^t x^{a_2} g^{b_2}\, dx \leqq u , \int_0^t x^{a_3} g^{b_3}\, dx \leqq v , \qquad g(x) \geqq 0 .$$

See

5. BELLMAN, R.: Adaptive control processes: a guided tour. Princeton, N. J.: Princeton University Press 1961.

Name Index

Subject Index