CMS Books in Mathematics

Constantin Niculescu
Lars-Erik Persson

Convex Functions and Their Applications

A Contemporary Approach

Canadian Mathematical Society
Société mathématique du Canada

CMS Books in Mathematics
Ouvrages de mathématiques de la SMC

1 Herman/Kučera/Šimša Equations and Inequalities

2 Arnold Abelian Groups and Representations of Finite Partially Ordered Sets

3 Borwein/Lewis Convex Analysis and Nonlinear Optimization

4 Levin/Lubinsky Orthogonal Polynomials for Exponential Weights

5 Kane Reflection Groups and Invariant Theory

6 Phillips Two Millennia of Mathematics

7 Deutsch Best Approximation in Inner Product Spaces

8 Fabian et al. Functional Analysis and Infinite-Dimensional Geometry

9 Křížek/Luca/Somer 17 Lectures on Fermat Numbers

10 Borwein Computational Excursions in Analysis and Number Theory

11 Reed/Sales (Editors) Recent Advances in Algorithms and Combinatorics

12 Herman/Kučera/Šimša Counting and Configurations

13 Nazareth Differentiable Optimization and Equation Solving

14 Phillips Interpolation and Approximation by Polynomials

15 Ben-Israel/Greville Generalized Inverses, Second Edition

16 Zhao Dynamical Systems in Population Biology

17 Göpfert et al. Variational Methods in Partially Ordered Spaces

18 Akivis/Goldberg Differential Geometry of Varieties with Degenerate Gauss Maps

19 Mikhalev/Shpilrain/Yu Combinatorial Methods

20 Borwein/Zhu Techniques of Variational Analysis

21 Van Brummelen/Kinyon Mathematics and the Historian's Craft: The Kenneth O. May Lectures

22 Lucchetti Convexity and Well-Posed Problems

23 Niculescu/Persson Convex Functions and Their Applications

24 Singer Duality for Nonconvex Approximation and Optimization

25 Higginson/Pimm/Sinclair Mathematics and the Aesthetic

Constantin P. Niculescu
Lars-Erik Persson

Convex Functions and Their Applications

A Contemporary Approach

With 8 Illustrations

Springer

Constantin P. Niculescu
Department of Mathematics
University of Craiova
Craiova 200585
Romania
cniculescu@central.ucv.ro

Lars-Erik Persson
Department of Mathematics
Luleå University of Technology
Luleå 97187
Sweden
larserik@sm.luth.se

Mathematics Subject Classification (2000): 26Axx, 26Bxx, 26Dxx, 46Axx, 52Axx

Library of Congress Control Number: 2005932044

ISBN-10: 0-387-24300-3
ISBN-13: 978-0387-24300-9

Printed on acid-free paper.

Printed in the United States of America. (MVY)

9 8 7 6 5 4 3 2 1

springeronline.com

To Liliana and Lena

Preface

It seems to me that the notion of convex function is just as fundamental as positive function or increasing function. If I am not mistaken in this, the notion ought to find its place in elementary expositions of the theory of real functions.

J. L. W. V. Jensen

Convexity is a simple and natural notion which can be traced back to Archimedes (circa 250 B.C.), in connection with his famous estimate of the value of π (by using inscribed and circumscribed regular polygons). He noticed the important fact that the perimeter of a convex figure is smaller than the perimeter of any other convex figure surrounding it.

As a matter of fact, we experience convexity all the time and in many ways. The most prosaic example is our upright position, which is secured as long as the vertical projection of our center of gravity lies inside the convex envelope of our feet. Also, convexity has a great impact on our everyday life through numerous applications in industry, business, medicine, and art. So do the problems of optimum allocation of resources and equilibrium of noncooperative games.

The theory of convex functions is part of the general subject of convexity, since a convex function is one whose epigraph is a convex set. Nonetheless it is an important theory per se, which touches almost all branches of mathematics. Graphical analysis is one of the first topics in mathematics which requires the concept of convexity. Calculus gives us a powerful tool in recognizing convexity, the second-derivative test. Miraculously, this has a natural generalization for the several variables case, the Hessian test. Motivated by some deep problems in optimization and control theory, convex function theory has been extended to the framework of infinite dimensional Banach spaces (and even further).

The recognition of the subject of convex functions as one that deserves to be studied in its own right is generally ascribed to J. L. W. V. Jensen [114], [115]. However he was not the first to deal with such functions. Among his predecessors we should recall here Ch. Hermite [102], O. Hölder [106] and O. Stolz [233]. During the twentieth century, there was intense research activity and significant results were obtained in geometric functional analysis, mathematical economics, convex analysis, and nonlinear optimization. A classic book by G. H. Hardy, J. E. Littlewood and G. Pólya [99] played a large role in the popularization of the subject of convex functions.

Roughly speaking, there are two basic properties of convex functions that make them so widely used in theoretical and applied mathematics:

- *The maximum is attained at a boundary point.*
- *Any local minimum is a global one. Moreover, a strictly convex function admits at most one minimum.*

The modern viewpoint on convex functions entails a powerful and elegant interaction between analysis and geometry. In a memorable paper dedicated to the Brunn–Minkowski inequality, R. J. Gardner [88, p. 358], described this reality in beautiful phrases: [convexity] "appears like an octopus, tentacles reaching far and wide, its shape and color changing as it roams from one area to the next. It is quite clear that research opportunities abound."

Over the years a number of notable books dedicated to the theory and applications of convex functions appeared. We mention here: L. Hörmander [108], M. A. Krasnosel'skii and Ya. B. Rutickii [132], J. E. Pečarić, F. Proschan and Y. C. Tong [196], R. R. Phelps [199], [200] and A. W. Roberts and D. E. Varberg [212]. The references at the end of this book include many other fine books dedicated to one aspect or another of the theory.

The title of the book by L. Hörmander, *Notions of Convexity,* is very suggestive for the present state of art. In fact, nowadays the study of convex functions has evolved into a larger theory about functions which are adapted to other geometries of the domain and/or obey other laws of comparison of means. Examples are log-convex functions, multiplicatively convex functions, subharmonic functions, and functions which are convex with respect to a subgroup of the linear group.

Our book aims to be a thorough introduction to contemporary convex function theory. It covers a large variety of subjects, from the one real variable case to the infinite dimensional case, including Jensen's inequality and its ramifications, the Hardy–Littlewood–Pólya theory of majorization, the theory of gamma and beta functions, the Borell–Brascamp–Lieb form of the Prékopa–Leindler inequality (as well as the connection with isoperimetric inequalities), Alexandrov's well-known result on the second differentiability of convex functions, the highlights of Choquet's theory, a brief account on the recent solution to Horn's conjecture, and many more. It is certainly a book where inequalities play a central role but in no case a book on inequalities. Many results are new, and the whole book reflects our own experiences, both in teaching and research.

This book may serve many purposes, ranging from a one-semester graduate course on Convex Functions and Applications to additional bibliographic material. In a course for first year graduate students, we used the following route:

- *Background*: Sections 1.1–1.3, 1.5, 1.7, 1.8, 1.10.
- *The beta and gamma functions*: Section 2.2.

- *Convex functions of several variables*: Sections 3.1–3.12.
- *The variational approach of partial differential equations*: Appendix C.

The necessary background is advanced calculus and linear algebra. This can be covered from many sources, for example, from *Analysis I* and *II* by S. Lang [137], [138]. A thorough presentation of the fundamentals of measure theory is also available in L. C. Evans and R. F. Gariepy [74]. For further reading we recommend the classical texts by F. H. Clarke [56] and I. Ekeland and R. Temam [70].

Our book is not meant to be read from cover to cover. For example, Section 1.9, which deals with the Hermite–Hadamard inequality, offers a good starting point for Choquet's theory. Then the reader may continue with Chapter 4, where this theory is presented in a slightly more general form, to allow the presence of certain signed measures. We recommend this chapter to be studied in parallel with the *Lectures on Choquet's theory* by R. R. Phelps [200]. For the reader's convenience, we collected in Appendix A all the necessary material on the separation of convex sets in locally convex Hausdorff spaces (as well as a proof of the Krein–Milman theorem).

Appendix B may be seen both as an illustration of convex function theory and an introduction to an important topic in real algebraic geometry: the theory of semi-algebraic sets.

Sections 3.11 and 3.12 offer all necessary background on a further study of convex geometric analysis, a fast-growing topic which relates many important branches of mathematics.

To help the reader in understanding the theory presented, each section ends with exercises (accompanied by hints). Also, each chapter ends with comments covering supplementary material and historical information. The primary sources we have relied upon for this book are listed in the references.

In order to avoid any confusion relative to our notation, a symbol index was added for the convenience of the reader. Notice that our book deals only with *real* linear spaces and all Borel measures under attention are assumed to be *regular*.

We wish to thank all our colleagues and friends who read and commented on various versions and parts of the manuscript: Madalina Deaconu, Andaluzia Matei, Sorin Micu, Florin Popovici, Mircea Preda, Thomas Strömberg, Andrei Vernescu, Peter Wall, Anna Wedestig and Tudor Zamfirescu.

We also acknowledge the financial support of Wenner–Gren Foundations (Grant 25 12 2002), which made possible the cooperation of the two authors.

In order to keep in touch with our readers, a web page for this book will be made available at http://www.inf.ucv.ro/~niculescu/Convex_Functions.html

Craiova and Luleå
September 2004

Constantin P. Niculescu
Lars-Erik Persson

Contents

List of symbols

$\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$: the classical numerical sets (naturals, integers etc.)
$\mathbb{N}^\star$: the set of positive integers
$\mathbb{R}_+$: the set of nonnegative real numbers
$\mathbb{R}_+^\star$: the set of positive real numbers
$\overline{\mathbb{R}}$: the set of extended real numbers
$\emptyset$: empty set

∂A: boundary of A
$\overline{A}$: closure of A
$\operatorname{int} A$: interior of A
A°: polar of A
$\operatorname{rbd}(A)$: relative boundary of A
$\operatorname{ri}(A)$: relative interior of A
$B_r(a)$: open ball center a, radius r
$\overline{B}_r(a)$: closed ball center a, radius r
$[x, y]$: line segment
$\operatorname{aff}(A)$: affine hull of A
$\operatorname{co}(A)$: convex hull of A
$\overline{\operatorname{co}}(A)$: closed convex hull of A
$\operatorname{Ext} K$: the extreme points of K
$\lambda A + \mu B = \{\lambda x + \mu y \mid x \in A, y \in B\}$
$|A|$: cardinality of A
$\operatorname{diam}(A)$: diameter of A
$\operatorname{Vol}_n(K)$: n-dimensional volume

$\mathcal{P}_C(x)$: set of best approximation from x

χ_A: characteristic function of A
id: identity
δ_C: indicator function
$f|_K$: restriction of f to K
f^*: (Legendre) conjugate function
$\underline{f}$: lower envelope of f
$\overline{f}$: upper envelope of f
$f^{\downarrow}$: symmetric-decreasing rearrangement of f
$d_U(x) = d(x, U)$: distance from x to U
$P_K(x)$: orthogonal projection
$\operatorname{dom}(f)$: effective domain of f
$\operatorname{epi}(f)$: epigraph of f
$\operatorname{graph}(f)$: graph of f
$\partial f(a)$: subdifferential of f at a
$\operatorname{supp}(f)$: support of f
$f * g$: convolution
$f \odot g$: infimal convolution

$\mathbb{R}^n$: Euclidean n-space
$\mathbb{R}^n_+ = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_1, \ldots, x_n \geq 0\}$, the nonnegative orthant
$\mathbb{R}^n_{++} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_1, \ldots, x_n > 0\}$
$\mathbb{R}^n_{\geq} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_1 \geq \cdots \geq x_n\}$
$\langle x, y \rangle$, $x \cdot y$: inner product
$\mathrm{M}_n(\mathbb{R})$, $\mathrm{M}_n(\mathbb{C})$: spaces of $n \times n$-dimensional matrices
$\mathrm{GL}(n, \mathbb{R})$: the group of nonsingular matrices
$\mathrm{Sym}^+(n, \mathbb{R})$: the set of all positive matrices of $\mathrm{M}_n(\mathbb{R})$
$\mathrm{Sym}^{++}(n, \mathbb{R})$: the set of all strictly positive matrices of $\mathrm{M}_n(\mathbb{R})$
$\dim E$: dimension of E
E': dual space
$A^\star$: adjoint matrix
$\det A$: determinant of A
$\ker A$: kernel (null space) of A
$\operatorname{rng} A$: range of A
$\operatorname{trace} A$: trace of A

$\liminf\limits_{x \to a} f(x) = \lim_{r \to 0} \inf\{f(x) \mid x \in \mathrm{dom}(f) \cap B_r(a)\}$: lower limit

$\limsup\limits_{x \to a} f(x) = \lim_{r \to 0} \sup\{f(x) \mid x \in \mathrm{dom}(f) \cap B_r(a)\}$: upper limit

$\underline{D}f(a)$ and $\overline{D}f(a)$: lower and upper derivatives

$\underline{\mathcal{D}}^2 f(a)$ and $\overline{\mathcal{D}}^2 f(a)$: lower and upper second symmetric derivatives

$f'_+(a; v)$ and $f'_-(a; v)$: lateral directional derivatives

$f'(a; v)$: first Gâteaux differential

$f''(a; v, w)$: second Gâteaux differential

df: first order differential

$d^2 f$: second order differential

$\dfrac{\partial f}{\partial x_k}$: partial derivative

$D^\alpha f = \dfrac{\partial^{\alpha_1 + \cdots + \alpha_n} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}$

∇: gradient

$\mathrm{Hess}_A f$: Hessian matrix of f at a

$\nabla^2 f(a)$: Alexandrov Hessian of f at a

$A(K)$: space of real-valued continuous and affine functions

$\mathrm{Conv}(K)$: space of real-valued continuous and convex functions

$C(K)$: space of real-valued continuous functions

$C^m(\Omega) = \{f \mid D^\alpha f \in C(\Omega) \text{ for all } |\alpha| \le m\}$

$C^m(\overline{\Omega}) = \{f \mid D^\alpha f \text{ uniformly continuous on } \Omega \text{ for all } |\alpha| \le m\}$

$C_c^\infty(\Omega)$: space of functions of class C^∞ with compact support

$L^p(\Omega)$: space of p-th-power Lebesgue integrable functions on Ω

$L^p(\mu)$: space of p-th-power μ-integrable functions

$\|f\|_{L^p}$: L^p-norm

$\operatorname{ess\,sup}$: essential supremum

ℓ^p: space of p-th-power absolutely summing real sequences

$\mathrm{Lip}(f)$: Lipschitz constant

$H^m(\Omega)$: Sobolev space on Ω

$\|f\|_{H^m}$: Sobolev norm

$H_0^m(\Omega)$: norm closure of $C_c^\infty(\Omega)$ in $H^m(\Omega)$

$\mathrm{Prob}(X)$: set of Borel probability measures on X

δ_a: Dirac measure concentrated at a

x_μ: barycenter of μ

$\mathcal{E}(f)$: conditional expectation of f

$A(s,t)$, $G(s,t)$, $H(s,t)$: arithmetic, geometric and harmonic means
$I(s,t)$: identric mean
$L(s,t)$: logarithmic mean
$M_p(s,t)$, $M_p(f;\mu)$: Hölder (power) mean
$M_{[\varphi]}$: quasi-arithmetic mean

Introduction

At the core of the notion of convexity is the comparison of means. By a *mean* (on an interval I) we understand any function $M\colon I \times I \to I$ which verifies the following property of *intermediacy*,

$$\inf\{s,t\} \leq M(s,t) \leq \sup\{s,t\},$$

for all pairs $\{s,t\}$ of elements of I. M is called a *strict* mean if these inequalities are strict for $s \neq t$, and M is called a *symmetric* mean if $M(s,t) = M(t,s)$ for all $s, t \in I$.

When I is one of the intervals $(0,\infty)$, $[0,\infty)$ or $(-\infty,\infty)$, it is usual to consider *homogeneous* means, that is,

$$M(\alpha s, \alpha t) = \alpha M(s,t)$$

for all $\alpha > 0$ and all $s, t \in I$.

Several examples of strict, symmetric and homogeneous means of strictly positive variables are listed below. They are all *continuous* (that is, continuous in both arguments).

Hölder's means (also called *power means*):

$$M_p(s,t) = ((s^p + t^p)/2)^{1/p}, \quad \text{for } p \neq 0$$
$$M_0(s,t) = \lim_{p\to 0} M_p(s,t) = \sqrt{st},$$

to which we can add

$$M_{-\infty}(s,t) = \inf\{s,t\} \quad \text{and} \quad M_{\infty}(s,t) = \sup\{s,t\}.$$

Then $A = M_1$ is the *arithmetic mean* and $G = M_0$ is the *geometric mean*. The mean $H = M_{-1}$ is known as the *harmonic mean*.

Lehmer's means:

$$L_p(s,t) = (s^p + t^p)/(s^{p-1} + t^{p-1}).$$

Notice that $L_1 = A$, $L_{1/2} = G$ and $L_0 = H$. These are the only means that are both Lehmer means and Hölder means.

Stolarsky's means:

$$S_p(s,t) = [(s^p - t^p)/(ps - pt)]^{1/(p-1)}, \quad p \neq 0, 1.$$

The limiting cases ($p = 0$ and $p = 1$) give the *logarithmic* and *identric* means, respectively. Thus

$$S_0(s,t) = \lim_{p\to 0} S_p(s,t) = \frac{s-t}{\log s - \log t} = L(s,t)$$
$$S_1(s,t) = \lim_{p\to 1} S_p(s,t) = \frac{1}{e}\Big(\frac{t^t}{s^s}\Big)^{1/(t-s)} = I(s,t).$$

Notice that $S_2 = A$ and $S_{-1} = G$. The reader may find a comprehensive account on the entire topic of means in [44].

An important mathematical problem is to investigate how functions behave under the action of means. The best-known case is that of *midpoint convex* (or *Jensen convex*) functions, which deal with the arithmetic mean. They are precisely the functions $f\colon I \to \mathbb{R}$ such that

$$f\Big(\frac{x+y}{2}\Big) \leq \frac{f(x)+f(y)}{2} \tag{J}$$

for all $x, y \in I$. In the context of continuity (which appears to be the only one of real interest), midpoint convexity means *convexity*, that is,

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y) \tag{C}$$

for all $x, y \in I$ and all $\lambda \in [0,1]$. See Theorem 1.1.4 for details. By mathematical induction we can extend the inequality (C) to the convex combinations of finitely many points in I and next to random variables associated to arbitrary probability spaces. These extensions are known as the *discrete Jensen inequality* and the *integral Jensen inequality*, respectively.

It turns out that similar results work when the arithmetic mean is replaced by any other mean with nice properties. For example, this is the case for regular means. A mean $M\colon I \times I \to \mathbb{R}$ is called *regular* if it is homogeneous, symmetric, continuous and also increasing in each variable (when the other is fixed). Notice that the Hölder means and the Stolarsky means are regular. The Lehmer's mean L_2 is not increasing (and thus it is not regular).

The regular means can be extended from pairs of real numbers to random variables associated to probability spaces through a process providing a nonlinear theory of integration.

Consider first the case of a discrete probability space (X, Σ, μ), where $X = \{1,2\}$, $\Sigma = \mathcal{P}(\{1,2\})$ and $\mu\colon \mathcal{P}(\{1,2\}) \to [0,1]$ is the probability measure such that $\mu(\{i\}) = \lambda_i$ for $i = 1,2$. A random variable associated to this space which takes values in I is a function

$$h\colon \{1,2\} \to I, \quad h(i) = x_i.$$

The mean M extends to a function $M(h\,;\mu) = M(x_1, x_2\,; \lambda_1, \lambda_2)$ such that $\inf h \leq M(h;\mu) \leq \sup h$ for all such random variables h. In this respect $M(x_1, x_2\,; \lambda_1, \lambda_2)$ appears as a weighted mean of x_1 and x_2 with weights λ_1 and λ_2 respectively. More precisely, we set

$$\begin{aligned} M(x_1, x_2\,; 1, 0) &= x_1 \\ M(x_1, x_2\,; 0, 1) &= x_2 \\ M(x_1, x_2\,; 1/2, 1/2) &= M(x_1, x_2) \end{aligned}$$

and for the other dyadic values of λ_1 and λ_2 we use formulas like

$$\begin{aligned} M(x_1, x_2\,; 3/4, 1/4) &= M(M(x_1, x_2), x_1) \\ M(x_1, x_2\,; 1/4, 3/4) &= M(M(x_1, x_2), x_2) \end{aligned}$$

and so on. In the general case, every $\lambda_1 \in (0,1)$, has a unique dyadic representation $\lambda_1 = \sum_{k=1}^{\infty} d_k/2^k$ (where $d_1, d_2, d_3, \ldots$ is a sequence consisting of 0's and 1's, which is not eventually 1) and we put

$$M(x_1, x_2\,; \lambda_1, \lambda_2) = \lim_{n\to\infty} M\Big(x_1, x_2\,; \sum_{k=1}^{n} d_k/2^k, 1 - \sum_{k=1}^{n} d_k/2^k\Big).$$

Now we can pass to the case of discrete probability spaces built on fields with three atoms via the formula

$$M(x_1, x_2, x_3\,; \lambda_1, \lambda_2, \lambda_3) = M\Big(M\Big(x_1, x_2\,; \frac{\lambda_1}{1-\lambda_3}, \frac{\lambda_2}{1-\lambda_3}\Big), x_3\,; 1-\lambda_3, \lambda_3\Big).$$

In the same manner, we can define the means $M(x_1, \ldots, x_n\,; \lambda_1, \ldots, \lambda_n)$, associated to random variables on probability spaces having n atoms.

We can bring together all power means M_p, for $p \in \mathbb{R}$, by considering the so called *quasi-arithmetic means*,

$$M_{[\varphi]}(s,t) = \varphi^{-1}\Big(\frac{1}{2}\varphi(s) + \frac{1}{2}\varphi(t)\Big),$$

which are associated to strictly monotone continuous mappings $\varphi\colon I \to \mathbb{R}$; the power mean M_p corresponds to $\varphi(x) = x^p$, if $p \neq 0$, and to $\varphi(x) = \log x$, if $p = 0$. For these means,

$$M_{[\varphi]}(x_1, \ldots, x_n\,; \lambda_1, \ldots, \lambda_n) = \varphi^{-1}\Big(\sum_{k=1}^{n} \lambda_k \varphi(x_k)\Big).$$

Particularly,

$$A(x_1, \ldots, x_n\,; \lambda_1, \ldots, \lambda_n) = \sum_{k=1}^{n} \lambda_k x_k,$$

in the case of the arithmetic mean, and

$$G(x_1,\ldots,x_n\,;\lambda_1,\ldots,\lambda_n)=\prod_{k=1}^{n}x_k^{\lambda_k},$$

in the case of the geometric mean.

The algorithm described above may lead to very complicated formulas for the weighted means $M(x_1,\ldots,x_n\,;\lambda_1,\ldots,\lambda_n)$ when M is not a quasi-arithmetic mean. For example, this is the case when M is the logarithmic mean L. However, the weighted means $L(x_1,\ldots,x_n\,;\lambda_1,\ldots,\lambda_n)$ can be introduced by a different algorithm proposed by A. O. Pittenger [201].

We can build a generalized theory of convexity (referred to as the theory of *comparative convexity*) simply, by replacing the arithmetic mean by other means. To be more specific, suppose there are given a pair of means M and N on the intervals I and J. A function $f\colon I\to J$ is called (M,N)-*midpoint affine*, (M,N)-*midpoint convex* and (M,N)-*midpoint concave* if, respectively,

$$\begin{aligned}f(M(x,y))&=N(f(x),f(y))\\f(M(x,y))&\leq N(f(x),f(y))\\f(M(x,y))&\geq N(f(x),f(y))\end{aligned}$$

for all $x,y\in I$ (see G. Aumann [13]). The condition of midpoint affinity is essentially a functional equation and this explains why the theory of comparative convexity has much in common with the subject of functional equations.

While the general theory of comparative convexity is still in its infancy, there are some notable facts to be mentioned here. For example, an easy inductive argument leads us to the following result:

Theorem A (The discrete form of Jensen's inequality) *If M and N are regular means and $F\colon I\to J$ is an (M,N)-midpoint convex continuous function, then*

$$F(M(x_1,\ldots,x_n\,;\lambda_1,\ldots,\lambda_n))\leq N((F(x_1),\ldots,F(x_n)\,;\lambda_1,\ldots,\lambda_n))$$

for all $x_1,\ldots,x_n\in I$ and all $\lambda_1,\ldots,\lambda_n\in[0,1]$ with $\sum_{k=1}^{n}\lambda_k=1$.

If (X,Σ,μ) is an arbitrary probability space, it is still possible to define the mean $M(h\,;\mu)$ for certain real random variables $h\in L^1(\mu)$ with values in I. In fact, letting $(\Sigma_\alpha)_\alpha$ be an upward directed net of finite subfields of Σ whose union generates Σ, the conditional expectation $\mathcal{E}(F|\Sigma_\alpha)$ of $F\in L^1(\mu)$ with respect to Σ_α gives rise to a positive contractive projection

$$P_\alpha\colon L^1(\mu)\to L^1(\mu|_{\Sigma_\alpha}),\quad P_\alpha(F)=\mathcal{E}(F|\Sigma_\alpha),$$

and

$$\mathcal{E}(F|\Sigma_\alpha)\to F\text{ in the norm topology of }L^1(\mu),$$

by Lebesgue's theorem on dominated convergence. See [103, p. 369].

A real random variable $h \in L^1(\mu)$ (with values in I) will be called *M-integrable* provided that the limit

$$M(h\,;\mu) = \lim_{\alpha} M(P_\alpha(h)\,;\mu|_{\Sigma_\alpha})$$

exists whenever $(\Sigma_\alpha)_\alpha$ is an upward directed net of finite subfields of Σ whose union generates Σ.

For the quasi-arithmetic mean $M_{[\varphi]}$ (associated to a strictly monotone continuous mapping $\varphi\colon I \to \mathbb{R}$) and the probability space associated to the restriction of the Lebesgue measure to an interval $[s,t] \subset I$, the construction above yields

$$M_{[\varphi]}\Big(\mathrm{id}_{[s,t]}\,;\frac{1}{t-s}dx\Big) = \varphi^{-1}\Big(\frac{1}{t-s}\int_s^t \varphi(x)\,dx\Big),$$

which coincides with the so-called *integral φ-mean* of s and t (also denoted $Int_\varphi(s,t)$). Using the fundamental theorem of calculus, it is easy to see that, on each interval I, the set of all integral means equals the set of all differential means. The *differential ψ-mean* of s and t (associated to a differentiable mapping $\psi\colon I \to \mathbb{R}$ for which ψ' is one-to-one) is given by the formula

$$D_\psi(s,t) = (\psi')^{-1}\Big(\frac{\psi(t)-\psi(s)}{t-s}\Big).$$

Passing to the limit in Theorem A we obtain:

Theorem B (The continuous form of Jensen's inequality) *Under the assumptions of Theorem* A, *if (X,Σ,μ) is a probability space, then*

$$F(M(h\,;\mu)) \le N((F\circ h\,;\mu))$$

for all $h \in L^1_{\mathbb{R}}(\mu)$ such that h is M-integrable and $F\circ h$ is N-integrable.

Theorem C (The Hermite–Hadamard inequality) *Suppose that M and N are regular means and $F\colon I \to J$ is a continuous function. Then F is (M,N)-midpoint convex if and only if for all $s < t$ in I and all probability measures μ on $[s,t]$ we have the inequality*

$$F(M(s\,;t)) \le N((F|_{[s,t]}\,;\mu)).$$

Proof. The necessity follows from Theorem B (applied to $h = \mathrm{id}_{[s,t]}$). The sufficiency represents the particular case where $\mu = (\delta_s + \delta_t)/2$. Here δ_x represents the Dirac measure concentrated at x. $\square$

It is worth mentioning the possibility of extending Theorem B beyond the class of probability measures. This can be done under the additional assumption of positive homogeneity (both for the means M and N, and the involved function F) following the model of Lebesgue theory, where formulæ such as

$$\int_{\mathbb{R}} f(x)\,dx = \lim_{n\to\infty}\left[2n\left(\frac{1}{2n}\int_{-n}^{n} f(x)\,dx\right)\right]$$

hold. Given a σ-finite measure space (X, Σ, μ), a function $h\colon X \to \mathbb{R}$ will be called *M-integrable* provided the limit

$$M(h\,;\mu) = \lim_{n\to\infty}\left[\mu(\Omega_n)\cdot M\left(h\chi_{\Omega_n}\,;\frac{\mu}{\mu(\Omega_n)}\right)\right]$$

exists for every increasing sequence $(\Omega_n)_n$ of finite measure sets of Σ with $\bigcup_n \Omega_n = X$. Then

$$F(M(h\,;\mu)) \le N((F\circ h\,;\mu))$$

for every $h \in L^1_{\mathbb{R}}(\mu)$ such that h is M-integrable and $F\circ h$ is N-integrable (a fact which extends Theorem C). An illustration of this construction is offered in Section 3.6.

The theory of comparative convexity encompasses a large variety of classes of convex-like functions, including log-convex functions, p-convex functions, and quasi-convex functions. While it is good to understand what they have in common, it is of equal importance to look inside their own fields.

Chapter 1 is devoted to the case of convex functions on intervals. We find there a rich diversity of results with important applications and deep generalizations to the context of several variables.

Chapter 2 is a specific presentation of other classes of functions acting on intervals which verify a condition of (M, N)-convexity. A theory on relative convexity, built on the concept of convexity of a function with respect to another function, is also included.

The basic theory of convex functions defined on convex sets in a normed linear space is presented in Chapter 3. The case of functions of several real variables offers many opportunities to illustrate the depth of the subject of convex functions through a number of powerful results: the existence of the orthogonal projection, the subdifferential calculus, the well-known Prékopa–Leindler inequality (and some of its ramifications), Alexandrov's beautiful result on the twice differentiability almost everywhere of a convex function, and the solution to the convex programming problem, among others.

Chapter 4 is devoted to Choquet's theory and its extension to the context of Steffensen–Popoviciu measures. This encompasses several remarkable results such as the Hermite–Hadamard inequality, the Jensen–Steffensen inequality, and Choquet's theorem on the existence of extremal measures.

As the material on convex functions (and their generalizations) is extremely vast, we had to restrict ourselves to some basic questions, leaving untouched many subjects which other people will probably consider of utmost importance. The Comments section at the end of each chapter, and the Appendices at the end of this book include many results and references to help the reader to get a better understanding of the field of convex functions.

1

Convex Functions on Intervals

The study of convex functions begins in the context of real-valued functions of a real variable. Here we find a rich variety of results with significant applications. More importantly, they will serve as a model for deep generalizations in the setting of several variables.

1.1 Convex Functions at First Glance

Throughout this book I will denote a nondegenerate interval.

Definition 1.1.1 A function $f\colon I \to \mathbb{R}$ is called *convex* if

$$f((1-\lambda)x + \lambda y) \le (1-\lambda)f(x) + \lambda f(y) \tag{1.1}$$

for all points x and y in I and all $\lambda \in [0,1]$. It is called *strictly convex* if the inequality (1.1) holds strictly whenever x and y are distinct points and $\lambda \in (0,1)$. If $-f$ is convex (respectively, strictly convex) then we say that f is *concave* (respectively, *strictly concave*). If f is both convex and concave, then f is said to be *affine*.

The affine functions on intervals are precisely the functions of the form $mx + n$, for suitable constants m and n. One can easily prove that the following three functions are convex (though not strictly convex): the positive part x^+, the negative part x^-, and the absolute value $|x|$. Together with the affine functions they provide the building blocks for the entire class of convex functions on intervals. See Theorem 1.5.7.

The convexity of a function $f\colon I \to \mathbb{R}$ means geometrically that the points of the graph of $f|_{[u,v]}$ are under (or on) the chord joining the endpoints $(u, f(u))$ and $(v, f(v))$, for all $u, v \in I$, $u < v$; see Fig. 1.1. Then

$$f(x) \le f(u) + \frac{f(v) - f(u)}{v - u}(x - u) \tag{1.2}$$

for all $x \in [u, v]$, and all $u, v \in I$, $u < v$. This shows that the convex functions are locally (that is, on any compact subinterval) majorized by affine functions. A companion result, concerning the existence of lines of support, will be presented in Section 1.5.

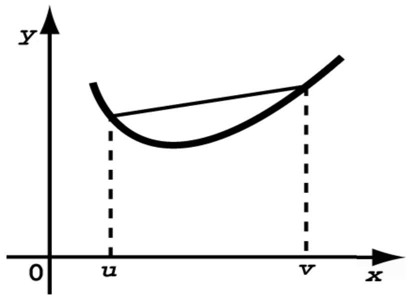

Fig. 1.1. Convex function: the graph is under the chord.

The intervals are closed under arbitrary convex combinations, that is,

$$\sum_{k=1}^{n} \lambda_k x_k \in I$$

for all $x_1, \ldots, x_n \in I$, and all $\lambda_1, \ldots, \lambda_n \in [0, 1]$ with $\sum_{k=1}^{n} \lambda_k = 1$. This can be proved by induction on the number n of points involved in the convex combinations. The case $n = 1$ is trivial, while for $n = 2$ it follows from the definition of a convex set. Assuming the result is true for all convex combinations with at most $n \geq 2$ points, let us pass to the case of combinations with $n+1$ points, $x = \sum_{k=1}^{n+1} \lambda_k x_k$. The nontrivial case is when all coefficients λ_k lie in $(0, 1)$. But in this case, due to our induction hypothesis, x can be represented as a convex combination of two elements of I,

$$x = (1 - \lambda_{n+1}) \Big(\sum_{k=1}^{n} \frac{\lambda_k}{1 - \lambda_{n+1}} x_k \Big) + \lambda_{n+1} x_{n+1},$$

hence x belongs to I.

The above remark has a notable counterpart for convex functions:

Lemma 1.1.2 (The discrete case of Jensen's inequality) *A real-valued function f defined on an interval I is convex if and only if for all $x_1, \ldots, x_n$ in I and all scalars $\lambda_1, \ldots, \lambda_n$ in $[0, 1]$ with $\sum_{k=1}^{n} \lambda_k = 1$ we have*

$$f \Big(\sum_{k=1}^{n} \lambda_k x_k \Big) \leq \sum_{k=1}^{n} \lambda_k f(x_k).$$

The above inequality is strict if f is strictly convex, all the points x_k are distinct and all scalars λ_k are positive.

A nice mechanical interpretation of this result was proposed by T. Needham [174]. The precision of Jensen's inequality is discussed in Section 1.4. See also Exercise 7, at the end of Section 1.8.

Related to the above geometrical interpretation of convexity is the following result due to S. Saks [219]:

Theorem 1.1.3 *Let f be a real-valued function defined on an interval I. Then f is convex if and only if for every compact subinterval J of I, and every affine function L, the supremum of $f + L$ on J is attained at an endpoint.*

This statement remains valid if the perturbations L are supposed to be linear (that is, of the form $L(x) = mx$ for suitable $m \in \mathbb{R}$).

Proof. Necessity: If f is convex, so is the sum $F = f + L$. Since every point of a subinterval $J = [x, y]$ is a convex combination $z = (1 - \lambda)x + \lambda y$ of x and y, we have

$$\begin{aligned} \sup_{z\in J} F(z) &= \sup_{\lambda\in[0,1]} F((1-\lambda)x + \lambda y) \\ &\leq \sup_{\lambda\in[0,1]} [(1-\lambda)F(x) + \lambda F(y)] = \max\{F(x), F(y)\}. \end{aligned}$$

Sufficiency: Given a compact subinterval $J = [x, y]$ of I, there exists an affine function $L(x) = mx + n$ which agrees with f at the two endpoints x and y. Then

$$\sup_{\lambda\in[0,1]} [(f - L)((1-\lambda)x + \lambda y)] = 0,$$

which yields

$$\begin{aligned} 0 &\geq f((1-\lambda)x + \lambda y) - L((1-\lambda)x + \lambda y) \\ &= f((1-\lambda)x + \lambda y) - (1-\lambda)L(x) - \lambda L(y) \\ &= f((1-\lambda)x + \lambda y) - (1-\lambda)f(x) - \lambda f(y) \end{aligned}$$

for every $\lambda \in [0, 1]$. $\square$

An easy consequence of Theorem 1.1.3 is that a convex function f is bounded on every compact subinterval $[u, v]$ of its interval of definition. In fact, $f(x) \leq M = \max\{f(u), f(v)\}$ on $[u, v]$ and writing an arbitrary point $x \in [u, v]$ in the form $x = (u + v)/2 + t$ for some t with $|t| \leq (v - u)/2$, we easily infer that

$$\begin{aligned} f(x) &= f\Big(\frac{u+v}{2} + t\Big) \geq 2f\Big(\frac{u+v}{2}\Big) - f\Big(\frac{u+v}{2} - t\Big) \\ &\geq 2f\Big(\frac{u+v}{2}\Big) - M. \end{aligned}$$

Checking that a function is convex or not is not very easy, but fortunately several useful criteria are available. Probably the simplest one is the following:

Theorem 1.1.4 (J. L. W. V. Jensen [115]) *Let $f\colon I \to \mathbb{R}$ be a continuous function. Then f is convex if and only if f is midpoint convex, that is,*

$$f\Big(\frac{x+y}{2}\Big) \leq \frac{f(x)+f(y)}{2} \quad \textit{for all} \;\; x, y \in I.$$

Proof. Clearly, only the sufficiency part needs an argument. By *reductio ad absurdum*, if f is not convex, then there exists a subinterval $[a, b]$ such that the graph of $f|_{[a,b]}$ is not under the chord joining $(a, f(a))$ and $(b, f(b))$; that is, the function

$$\varphi(x) = f(x) - \frac{f(b)-f(a)}{b-a}(x-a) - f(a), \quad x \in [a, b]$$

verifies $\gamma = \sup\{\varphi(x) \mid x \in [a, b]\} > 0$. Notice that φ is continuous and $\varphi(a) = \varphi(b) = 0$. Also, a direct computation shows that φ is also midpoint convex. Put $c = \inf\{x \in [a, b] \mid \varphi(x) = \gamma\}$; then necessarily $\varphi(c) = \gamma$ and $c \in (a, b)$. By the definition of c, for every $h > 0$ for which $c \pm h \in (a, b)$ we have

$$\varphi(c-h) < \varphi(c) \quad \text{and} \quad \varphi(c+h) \leq \varphi(c)$$

so that

$$\varphi(c) > \frac{\varphi(c-h) + \varphi(c+h)}{2}$$

in contradiction with the fact that φ is midpoint convex. $\square$

Corollary 1.1.5 *Let $f\colon I \to \mathbb{R}$ be a continuous function. Then f is convex if and only if*

$$f(x+h) + f(x-h) - 2f(x) \geq 0$$

for all $x \in I$ and all $h > 0$ such that both $x+h$ and $x-h$ are in I.

Notice that both Theorem 1.1.4 and its Corollary 1.1.5 above have straightforward variants for the case of strictly convex functions.

Corollary 1.1.5 allows us to check immediately the strict convexity of some very common functions, such as the exponential function. Indeed, due to the fact that

$$a, b > 0,\ a \neq b, \quad \text{implies} \quad \frac{a+b}{2} > \sqrt{ab}$$

we have

$$\mathrm{e}^{x+h} + \mathrm{e}^{x-h} - 2\mathrm{e}^{x} > 0$$

for all $x \in \mathbb{R}$ and all $h > 0$. An immediate consequence of this remark is the following result, which extends the well-known arithmetic mean–geometric mean inequality (abbreviated, AM–GM inequality):

Theorem 1.1.6 (The weighted form of the AM–GM inequality; L. J. Rogers [215]) *If* $x_1, \dots, x_n \in (0, \infty)$ *and* $\lambda_1, \dots, \lambda_n \in (0, 1)$, $\sum_{k=1}^{n} \lambda_k = 1$, *then*

$$\sum_{k=1}^{n} \lambda_k x_k > x_1^{\lambda_1} \cdots x_n^{\lambda_n}$$

unless $x_1 = \cdots = x_n$.

Replacing x_k by $1/x_k$ in the last inequality we get (under the same hypotheses on x_k and λ_k),

$$x_1^{\lambda_1} \cdots x_n^{\lambda_n} > 1 \Big/ \sum_{k\,=\,1}^{n} \frac{\lambda_k}{x_k}$$

unless $x_1 = \cdots = x_n$ (which represents the weighted form of the *geometric mean–harmonic mean inequality*).

The particular case of Theorem 1.1.6 where $\lambda_1 = \cdots = \lambda_n = 1/n$ represents the usual *AM–GM inequality*, which can be completed as above, with its relation to the harmonic mean: *For every family* $x_1, \dots, x_n$ *of positive numbers we have*

$$\frac{x_1 + \cdots + x_n}{n} > \sqrt[n]{x_1 \cdots x_n} > \frac{n}{\left(\frac{1}{x_1} + \cdots + \frac{1}{x_n}\right)}$$

unless $x_1 = \cdots = x_n$. An estimate of these inequalities is the objective of Section 2.5 below.

The closure under functional operations with convex functions is an important source of examples in this area.

Proposition 1.1.7 (The operations with convex functions)

(i) *Adding two convex functions* (*defined on the same interval*) *we obtain a convex function; if one of them is strictly convex, then the sum is also strictly convex.*

(ii) *Multiplying a* (*strictly*) *convex function by a positive scalar we obtain also a* (*strictly*) *convex function.*

(iii) *The restriction of every* (*strictly*) *convex function to a subinterval of its domain is also a* (*strictly*) *convex function.*

(iv) *If* $f \colon I \to \mathbb{R}$ *is a convex* (*respectively a strictly convex*) *function and* $g \colon \mathbb{R} \to \mathbb{R}$ *is a nondecreasing* (*respectively an increasing*) *convex function, then* $g \circ f$ *is convex* (*respectively strictly convex*).

(v) *Suppose that* f *is a bijection between two intervals* I *and* J. *If* f *is increasing, then* f *is* (*strictly*) *convex if and only if* f^{-1} *is* (*strictly*) concave. *If* f *is a decreasing bijection, then* f *and* f^{-1} *are of the same type of convexity.*

We end this section with an analogue of Theorem 1.1.4 for triplets:

Theorem 1.1.8 (Popoviciu's inequality [206]) *Let $f\colon I \to \mathbb{R}$ be a continuous function. Then f is convex if and only if*

$$\frac{f(x)+f(y)+f(z)}{3} + f\Big(\frac{x+y+z}{3}\Big) \geq \frac{2}{3}\Big[f\Big(\frac{x+y}{2}\Big) + f\Big(\frac{y+z}{2}\Big) + f\Big(\frac{z+x}{2}\Big)\Big]$$

for all $x, y, z \in I$.

In the variant of strictly convex functions the above inequality is strict except for $x = y = z$.

Proof. Necessity: (this implication does not need the assumption on continuity). Without loss of generality we may assume that $x \leq y \leq z$. If $y \leq (x+y+z)/3$, then

$$(x+y+z)/3 \leq (x+z)/2 \leq z \quad \text{and} \quad (x+y+z)/3 \leq (y+z)/2 \leq z,$$

which yields two numbers $s, t \in [0,1]$ such that

$$\begin{aligned} \frac{x+z}{2} &= s \cdot \frac{x+y+z}{3} + (1-s) \cdot z \\ \frac{y+z}{2} &= t \cdot \frac{x+y+z}{3} + (1-t) \cdot z. \end{aligned}$$

Summing up, we get $(x+y-2z)(s+t-3/2) = 0$. If $x+y-2z = 0$, then necessarily $x = y = z$, and Popoviciu's inequality is clear.

If $s + t = 3/2$, we have to sum up the following three inequalities:

$$\begin{aligned} f\Big(\frac{x+z}{2}\Big) &\leq s \cdot f\Big(\frac{x+y+z}{3}\Big) + (1-s) \cdot f(z) \\ f\Big(\frac{y+z}{2}\Big) &\leq t \cdot f\Big(\frac{x+y+z}{3}\Big) + (1-t) \cdot f(z) \\ f\Big(\frac{x+y}{2}\Big) &\leq \frac{1}{2} \cdot f(x) + \frac{1}{2} \cdot f(y) \end{aligned}$$

and then multiply both sides by $2/3$.

The case where $(x+y+z)/3 < y$ can be treated in a similar way.

Sufficiency: Popoviciu's inequality (when applied for $y = z$), yields the following substitute for the condition of midpoint convexity:

$$\frac{1}{4} f(x) + \frac{3}{4} f\Big(\frac{x+2y}{3}\Big) \geq f\Big(\frac{x+y}{2}\Big) \quad \text{for all } x, y \in I. \tag{1.3}$$

Using this remark, the proof follows verbatim the argument of Theorem 1.1.4 above. $\square$

The above statement of Popoviciu's inequality is only a simplified version of a considerably more general result; see the Comments at the end of this chapter. However, even this version leads to interesting inequalities; see Exercise 9. An estimate from below of Popoviciu's inequality is available in [188].

Exercises

1. Prove that the following functions are strictly convex:

- $-\log x$ and $x \log x$ on $(0, \infty)$;
- x^p on $[0, \infty)$ if $p > 1$; x^p on $(0, \infty)$ if $p < 0$; $-x^p$ on $[0, \infty)$ if $p \in (0, 1)$;
- $(1 + x^p)^{1/p}$ on $[0, \infty)$ if $p > 1$.

2. Let $f \colon I \to \mathbb{R}$ be a convex function and let $x_1, \ldots, x_n \in I$ $(n \geq 2)$. Prove that

$$(n-1)\Big[\frac{f(x_1) + \cdots + f(x_{n-1})}{n-1} - f\Big(\frac{x_1 + \cdots + x_{n-1}}{n-1}\Big)\Big]$$

cannot exceed

$$n\Big[\frac{f(x_1) + \cdots + f(x_n)}{n} - f\Big(\frac{x_1 + \cdots + x_n}{n}\Big)\Big].$$

3. Let $x_1, \ldots, x_n > 0$ $(n \geq 2)$ and for each $1 \leq k \leq n$ put

$$A_k = \frac{x_1 + \cdots + x_k}{k} \quad \text{and} \quad G_k = (x_1 \cdots x_k)^{1/k}.$$

(i) (T. Popoviciu) Prove that

$$\Big(\frac{A_n}{G_n}\Big)^n \geq \Big(\frac{A_{n-1}}{G_{n-1}}\Big)^{n-1} \geq \cdots \geq \Big(\frac{A_1}{G_1}\Big)^1 = 1.$$

(ii) (R. Rado) Prove that

$$n(A_n - G_n) \geq (n-1)(A_{n-1} - G_{n-1}) \geq \cdots \geq 1 \cdot (A_1 - G_1) = 0.$$

[*Hint*: Apply the result of Exercise 2 to $f = -\log$ and respectively to $f = \exp$.]

4. Suppose that $f_1, \ldots, f_n$ are nonnegative concave functions with the same domain of definition. Prove that $(f_1 \cdots f_n)^{1/n}$ is also a concave function.

5. (i) Prove that Theorem 1.1.4 remains true if the condition of midpoint convexity is replaced by: $f((1-\alpha)x + \alpha y) \leq (1-\alpha)f(x) + \alpha f(y)$ for some fixed parameter $\alpha \in (0, 1)$, and for all $x, y \in I$.

(ii) Prove that Theorem 1.1.4 remains true if the condition of continuity is replaced by boundedness from above on every compact subinterval.

6. (New from old) Assume that $f(x)$ is a (strictly) convex function for $x > 0$. Prove that $xf(1/x)$ is (strictly) convex too.

7. Infer from Theorem 1.1.6 that $\min\limits_{x,y>0} (x + y + \frac{1}{x^2 y}) = 4/\sqrt{2}$.

8. (The power means in the discrete case: see Section 1.8, Exercise 1, for the integral case.) Let $x = (x_1, \ldots, x_n)$ and $\alpha = (\alpha_1, \ldots, \alpha_n)$ be two n-tuples of positive elements, such that $\sum_{k=1}^{n} \alpha_k = 1$. The (weighted) *power mean* of order t is defined as

$$M_t(x;\alpha) = \Big(\sum_{k=1}^{n} \alpha_k x_k^t\Big)^{1/t} \quad \text{for } t \neq 0$$

and

$$M_0(x;\alpha) = \lim_{t\to 0+} M_t(x,\alpha) = \prod_{k=1}^{n} x_k^{\alpha_k}.$$

Notice that M_1 is the arithmetic mean, M_0 is the geometric mean and M_{-1} is the harmonic mean. Moreover, $M_{-t}(x;\alpha) = M_t(x^{-1};\alpha)^{-1}$.

(i) Apply Jensen's inequality to the function $x^{t/s}$, to prove that

$$s \leq t \quad \text{implies} \quad M_s(x;\alpha) \leq M_t(x;\alpha).$$

(ii) Prove that the function $t \to t \log M_t(x;\alpha)$ is convex on $\mathbb{R}$.

(iii) We define $M_{-\infty}(x;\alpha) = \inf\{x_k \mid k\}$ and $M_\infty(x,\alpha) = \sup\{x_k \mid k\}$. Prove that

$$\lim_{t\to -\infty} M_t(x;\alpha) = M_{-\infty}(x;\alpha) \quad \text{and} \quad \lim_{t\to\infty} M_t(x;\alpha) = M_\infty(x;\alpha).$$

9. (An illustration of Popoviciu's inequality) Suppose that x_1, x_2, x_3 are positive numbers, not all equal. Prove that:

(i) $27\prod_{i<j}(x_i + x_j)^2 > 64x_1x_2x_3(x_1 + x_2 + x_3)^3$;

(ii) $x_1^6 + x_2^6 + x_3^6 + 3x_1^2x_2^2x_3^2 > 2(x_1^3x_2^3 + x_2^3x_3^3 + x_3^3x_1^3)$.

1.2 Young's Inequality and Its Consequences

Young's inequality asserts that

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q} \quad \text{for all } a, b \geq 0,$$

whenever $p, q \in (1,\infty)$ and $1/p + 1/q = 1$; the equality holds if and only if $a^p = b^q$. This is a consequence of the strict convexity of the exponential function. In fact,

$$\begin{aligned} ab &= \mathrm{e}^{\log ab} = \mathrm{e}^{(1/p)\log a^p + (1/q)\log b^q} \\ &< \frac{1}{p}\,\mathrm{e}^{\log a^p} + \frac{1}{q}\,\mathrm{e}^{\log b^q} = \frac{a^p}{p} + \frac{b^q}{q} \end{aligned}$$

for all $a, b > 0$ with $a^p \neq b^q$. An alternative argument can be obtained by studying the variation of the function

$$F(a) = \frac{a^p}{p} + \frac{b^q}{q} - ab, \quad a \ge 0,$$

where $b \ge 0$ is a parameter. Then F has a strict global minimum at $a = b^{q/p}$, which yields $F(a) > F(b^{q/p}) = 0$ for all $a \ge 0$, $a \ne b^{q/p}$.

W. H. Young [247] actually proved a much more general inequality which yields the aforementioned one for $f(x) = x^{p-1}$:

Theorem 1.2.1 (Young's inequality) *Suppose that* $f\colon [0,\infty) \to [0,\infty)$ *is an increasing continuous function such that* $f(0) = 0$ *and* $\lim_{x\to\infty} f(x) = \infty$. *Then*

$$ab \le \int_0^a f(x)\,dx + \int_0^b f^{-1}(x)\,dx$$

for all $a, b \ge 0$, *and equality occurs if and only if* $b = f(a)$.

Proof. Using the definition of the derivative we can easily prove that the function

$$F(x) = \int_0^x f(t)\,dt + \int_0^{f(x)} f^{-1}(t)\,dt - xf(x)$$

is differentiable, with F' identically 0. This yields

$$0 \le u \le a \text{ and } 0 \le v \le f(a) \implies uv \le \int_0^u f(t)\,dt + \int_0^v f^{-1}(t)\,dt$$

and the conclusion of the theorem is now clear. $\square$

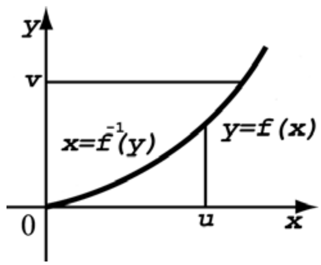

Fig. 1.2. The areas of the two curvilinear triangles exceed the area of the rectangle with sides u and v.

The geometric meaning of Young's inequality is indicated in Fig. 1.2.

Young's inequality is the source of many basic inequalities. The next two applications concern complex functions defined on an arbitrary measure space (X, Σ, μ).

Theorem 1.2.2 (The Rogers–Hölder inequality for $p > 1$) *Let $p, q \in (1, \infty)$ with $1/p + 1/q = 1$, and let $f \in L^p(\mu)$ and $g \in L^q(\mu)$. Then fg is in $L^1(\mu)$ and we have*

$$\left| \int_X fg \, d\mu \right| \leq \int_X |fg| \, d\mu \tag{1.4}$$

and

$$\int_X |fg| \, d\mu \leq \|f\|_{L^p} \, \|g\|_{L^q} \tag{1.5}$$

and thus

$$\left| \int_X fg \, d\mu \right| \leq \|f\|_{L^p} \, \|g\|_{L^q}. \tag{1.6}$$

The above result extends in a straightforward manner to the pairs $p = 1$, $q = \infty$ and $p = \infty$, $q = 1$. In the complementary domain, $p \in (-\infty, 1) \backslash \{0\}$ and $1/p + 1/q = 1$, the inequality sign in (1.4)–(1.6) should be reversed. See Exercises 3 and 4.

For $p = q = 2$, the inequality (1.6) is called the *Cauchy–Buniakovski–Schwarz inequality*.

Proof. The first inequality is trivial. If f or g is zero μ-almost everywhere, then the second inequality is trivial. Otherwise, using Young's inequality, we have

$$\frac{|f(x)|}{\|f\|_{L^p}} \cdot \frac{|g(x)|}{\|g\|_{L^q}} \leq \frac{1}{p} \cdot \frac{|f(x)|^p}{\|f\|_{L^p}^p} + \frac{1}{q} \cdot \frac{|g(x)|^q}{\|g\|_{L^q}^q}$$

for all x in X, such that $fg \in L^1(\mu)$. Thus

$$\frac{1}{\|f\|_{L^p} \|g\|_{L^q}} \int_X |fg| \, d\mu \leq 1$$

and this proves (1.5). The inequality (1.6) is immediate. □

Remark 1.2.3 (Conditions for equality in Theorem 1.2.2) The basic observation is the fact that

$$f \geq 0 \quad \text{and} \quad \int_X f \, d\mu = 0 \quad \text{imply} \quad f = 0 \ \mu\text{-almost everywhere.}$$

Consequently we have equality in (1.4) if and only if

$$f(x)g(x) = e^{i\theta} |f(x)g(x)|$$

for some real constant θ and for μ-almost every x.

Suppose that $p, q \in (1, \infty)$ and f and g are not zero μ-almost everywhere. In order to get equality in (1.5) it is necessary and sufficient to have

$$\frac{|f(x)|}{\|f\|_{L^p}} \cdot \frac{|g(x)|}{\|g\|_{L^q}} = \frac{1}{p} \cdot \frac{|f(x)|^p}{\|f\|_{L^p}^p} + \frac{1}{q} \cdot \frac{|g(x)|^q}{\|g\|_{L^q}^q}$$

almost everywhere. The equality case in Young's inequality shows that this is equivalent to $|f(x)|^p/\|f\|_{L^p}^p = |g(x)|^q/\|g\|_{L^q}^q$ almost everywhere, that is,

$$A|f(x)|^p = B|g(x)|^q \quad \text{almost everywhere}$$

for some nonnegative numbers A and B.

If $p = 1$ and $q = \infty$, we have equality in (1.5) if and only if there is a constant $\lambda \geq 0$ such that $|g(x)| \leq \lambda$ almost everywhere, and $|g(x)| = \lambda$ for almost every point where $f(x) \neq 0$.

Theorem 1.2.4 (Minkowski's inequality) *For* $1 \leq p < \infty$ *and* $f, g \in L^p(\mu)$ *we have*

$$\|f + g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p}. \tag{1.7}$$

In the discrete case, using the notation of Exercise 8 in Section 1.1, this inequality reads

$$M_p(x + y, \alpha) \leq M_p(x, \alpha) + M_p(y, \alpha). \tag{1.8}$$

In this form, it extends to the complementary range $0 < p < 1$, with the inequality sign reversed. The integral analogue for $p < 1$ is presented in Section 3.6.

Proof. For $p = 1$, the inequality (1.7) follows immediately by integrating the inequality $|f + g| \leq |f| + |g|$. For $p \in (1, \infty)$ we have

$$\begin{aligned} |f + g|^p &\leq (|f| + |g|)^p \leq (2 \sup\{|f|, |g|\})^p \\ &\leq 2^p(|f|^p + |g|^p), \end{aligned}$$

which shows that $f + g \in L^p(\mu)$. Moreover, according to Theorem 1.2.2,

$$\begin{aligned} \|f + g\|_{L^p}^p &= \int_X |f + g|^p \, d\mu \leq \int_X |f + g|^{p-1} \, |f| \, d\mu + \int_X |f + g|^{p-1} \, |g| \, d\mu \\ &\leq \left(\int_X |f|^p \, d\mu \right)^{1/p} \left(\int_X |f + g|^{(p-1)q} \, d\mu \right)^{1/q} \\ &\quad + \left(\int_X |g|^p \, d\mu \right)^{1/p} \left(\int_X |f + g|^{(p-1)q} \, d\mu \right)^{1/q} \\ &= (\|f\|_{L^p} + \|g\|_{L^p}) \|f + g\|_{L^p}^{p/q}, \end{aligned}$$

where $1/p + 1/q = 1$, and it remains to observe that $p - p/q = 1$. $\square$

Remark 1.2.5 If $p = 1$, we obtain equality in (1.7) if and only if there is a positive measurable function φ such that

$$f(x)\varphi(x) = g(x)$$

almost everywhere on the set $\{x \mid f(x)g(x) \neq 0\}$.

If $p \in (1, \infty)$ and f is not 0 almost everywhere, then we have equality in (1.7) if and only if $g = \lambda f$ almost everywhere, for some $\lambda \geq 0$.

In the particular case when (X, Σ, μ) is the measure space associated with the counting measure on a finite set,

$$\mu\colon \mathcal{P}(\{1, \ldots, n\}) \to \mathbb{N}, \quad \mu(A) = |A|,$$

we retrieve the classical discrete forms of the above inequalities. For example, the discrete version of the Rogers–Hölder inequality can be read

$$\Big|\sum_{k=1}^{n} \xi_k \eta_k\Big| \leq \Big(\sum_{k=1}^{n} |\xi_k|^p\Big)^{1/p} \Big(\sum_{k=1}^{n} |\eta_k|^q\Big)^{1/q}$$

for all ξ_k, $\eta_k \in \mathbb{C}$, $k \in \{1, \ldots, n\}$. On the other hand, a moment's reflection shows that we can pass immediately from these discrete inequalities to their integral analogues, corresponding to finite measure spaces.

Remark 1.2.6 It is important to notice that all numerical inequalities of the form

$$f(x_1, \ldots, x_n) \geq 0 \quad \text{for } x_1, \ldots, x_n \geq 0 \tag{1.9}$$

where f is a continuous and positively homogeneous function of degree 1 (that is, $f(\lambda x_1, \ldots, \lambda x_n) = \lambda f(x_1, \ldots, x_n)$ for $\lambda \geq 0$), extend to the context of Banach lattices, via a functional calculus invented by A. J. Yudin and J. L. Krivine. This allows us to replace the real variables of f by positive elements of a Banach lattice. See [147, Vol. 2, pp. 40–43]. Particularly, this is the case of the AM–GM inequality, Rogers–Hölder's inequality, and Minkowski's inequality.

Also, all numerical inequalities of the form (1.9), attached to continuous functions, extend (via the functional calculus with self-adjoint elements) to the context of C^*-algebras. In fact, the n-tuples of real numbers can be replaced by n-tuples of mutually commuting positive elements of a C^*-algebra. See [58].

Exercises

1. Recall the *identity of Lagrange*,

$$\Big(\sum_{k=1}^{n} a_k^2\Big)\Big(\sum_{k=1}^{n} b_k^2\Big) = \sum_{1 \leq j < k \leq n} (a_j b_k - a_k b_j)^2 + \Big(\sum_{k=1}^{n} a_k b_k\Big)^2,$$

which works for all $a_k, b_k \in \mathbb{C}$, $k \in \{1, \ldots, n\}$. Infer from it the discrete form of Cauchy–Buniakovski–Schwarz inequality,

$$\Big|\sum_{k=1}^{n} \xi_k \eta_k\Big| \leq \Big(\sum_{k=1}^{n} |\xi_k|^2\Big)^{1/2} \Big(\sum_{k=1}^{n} |\eta_k|^2\Big)^{1/2},$$

and settle the equality case.

2. (The Bernoulli inequality)

(i) Prove that for all $x > -1$ we have

$$(1+x)^\alpha \geq 1 + \alpha x \quad \text{if } \alpha \in (-\infty, 0] \cup [1, \infty)$$

and

$$(1+x)^\alpha \leq 1 + \alpha x \quad \text{if } \alpha \in [0, 1];$$

if $\alpha \notin \{0, 1\}$, the equality occurs only for $x = 0$.

(ii) The substitution $1 + x \to x/y$ followed by a multiplication by y leads us to Young's inequality (for full range of parameters). Show that this inequality can be written

$$xy \geq \frac{x^p}{p} + \frac{y^q}{q} \quad \text{for all } x, y > 0$$

in the domain $p \in (-\infty, 1) \backslash \{0\}$ and $1/p + 1/q = 1$.

3. (The Rogers–Hölder inequality for $p \in (-\infty, 1) \backslash \{0\}$ and $1/p + 1/q = 1$) Apply Young's inequality to prove that

$$\sum_{k=1}^n |a_k b_k| \geq \Big(\sum_{k=1}^n |a_k|^p\Big)^{1/p} \Big(\sum_{k=1}^n |b_k|^q\Big)^{1/q}$$

for all $a_1, \ldots, a_n$, $b_1, \ldots, b_n \in \mathbb{C}$ and all $n \in \mathbb{N}^*$.

4. (A symmetric form of Rogers–Hölder inequality) Let p, q, r be nonzero real numbers such that $1/p + 1/q = 1/r$.

(i) Prove that the inequality

$$\Big(\sum_{k=1}^n \lambda_k |a_k b_k|^r\Big)^{1/r} \leq \Big(\sum_{k=1}^n \lambda_k |a_k|^p\Big)^{1/p} \Big(\sum_{k=1}^n \lambda_k |b_k|^q\Big)^{1/q}$$

holds in each of the following three cases:

$$p > 0,\ q > 0,\ r > 0; \quad p < 0,\ q > 0,\ r < 0; \quad p > 0,\ q < 0,\ r < 0.$$

(ii) Prove that the opposite inequality holds in each of the following cases:

$$p > 0,\ q < 0,\ r > 0; \quad p < 0,\ q > 0,\ r > 0; \quad p < 0,\ q < 0,\ r < 0.$$

Here $\lambda_1, \ldots, \lambda_n > 0$, $\sum_{k=1}^n \lambda_k = 1$, and $a_1, \ldots, a_n$, $b_1, \ldots, b_n \in \mathbb{C} \backslash \{0\}$, $n \in \mathbb{N}^*$.

(iii) Formulate the above inequalities in terms of power means and then prove they still work for $r = pq/(p+q)$ if p and q are not both zero, and $r = 0$ if $p = q = 0$.

5. Prove the following generalization of the Rogers–Hölder inequality: If (X, Σ, μ) is a measure space and $f_1, \ldots, f_n$ are functions such that $f_k \in L^{p_k}(\mu)$ for some $p_k \geq 1$, and $\sum_{k=1}^n 1/p_k = 1$, then

$$\left| \int_X \Big(\prod_{k=1}^n f_k \Big)\, d\mu \right| \leq \prod_{k=1}^n \|f_k\|_{L^{p_k}}.$$

6. (A general form of Minkowski's inequality, see [144, p. 47]) Suppose that $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are two σ-finite measure spaces, f is a nonnegative function on $X \times Y$ which is $\mu \times \nu$-measurable, and let $p \in [1, \infty)$. Then

$$\left(\int_X \left(\int_Y f(x, y)\, d\nu(y) \right)^p d\mu(x) \right)^{1/p} \leq \int_Y \left(\int_X f(x, y)^p\, d\mu(x) \right)^{1/p} d\nu(y).$$

1.3 Smoothness Properties

The entire discussion on the smoothness properties of convex functions on intervals is based on their characterization in terms of slopes of variable chords through arbitrary fixed points of their graphs.

Given a function $f\colon I \to \mathbb{R}$ and a point $a \in I$, one can associate with them a new function,

$$s_a\colon I \backslash \{a\} \to \mathbb{R}, \quad s_a(x) = \frac{f(x) - f(a)}{x - a},$$

whose value at x is the slope of the chord joining the points $(a, f(a))$ and $(x, f(x))$ of the graph of f.

Theorem 1.3.1 (L. Galvani [86]) *Let f be a real function defined on an interval I. Then f is convex (respectively, strictly convex) if and only if the associated functions s_a are nondecreasing (respectively, increasing).*

In fact,

$$\frac{s_a(y) - s_a(x)}{y - x} = \begin{vmatrix} 1 & x & f(x) \\ 1 & y & f(y) \\ 1 & a & f(a) \end{vmatrix} \Bigg/ \begin{vmatrix} 1 & x & x^2 \\ 1 & y & y^2 \\ 1 & a & a^2 \end{vmatrix}$$

for all three distinct points a, x, y of I, and the proof of Theorem 1.3.1 is a consequence of the following lemma:

Lemma 1.3.2 *Let f be a real function defined on an interval I. Then f is convex if and only if*

$$\begin{vmatrix} 1 & x & f(x) \\ 1 & y & f(y) \\ 1 & z & f(z) \end{vmatrix} \Bigg/ \begin{vmatrix} 1 & x & x^2 \\ 1 & y & y^2 \\ 1 & z & z^2 \end{vmatrix} \geq 0$$

for all three distinct points x, y, z of I; equivalently, if and only if

$$\begin{vmatrix} 1 & x & f(x) \\ 1 & y & f(y) \\ 1 & z & f(z) \end{vmatrix} \geq 0 \tag{1.10}$$

for all $x < y < z$ in I.

The corresponding variant for strict convexity is valid too, provided that $\geq$ is replaced by $>$.

Proof. The condition (1.10) means that

$$(z-y)f(x) - (z-x)f(y) + (y-x)f(z) \geq 0$$

for all $x < y < z$ in I. Since each y between x and z can be written as $y = (1-\lambda)x + \lambda z$, the latter condition is equivalent to the assertion that

$$f((1-\lambda)x + \lambda z) \leq (1-\lambda)f(x) + \lambda f(z)$$

for all $x < z$ in I and all $\lambda \in [0, 1]$. □

We are now prepared to state the main result on the smoothness of convex functions.

Theorem 1.3.3 (O. Stolz [233]) *Let $f\colon I \to \mathbb{R}$ be a convex function. Then f is continuous on the interior* $\operatorname{int} I$ *of I and has finite left and right derivatives at each point of* $\operatorname{int} I$. *Moreover, $x < y$ in* $\operatorname{int} I$ *implies*

$$f'_-(x) \leq f'_+(x) \leq f'_-(y) \leq f'_+(y).$$

Particularly, both f'_- and f'_+ are nondecreasing on $\operatorname{int} I$.

Proof. In fact, according to Theorem 1.3.1 above, we have

$$\frac{f(x)-f(a)}{x-a} \leq \frac{f(y)-f(a)}{y-a} \leq \frac{f(z)-f(a)}{z-a}$$

for all $x \leq y < a < z$ in I. This fact assures us that the left derivative at a exists and

$$f'_-(a) \leq \frac{f(z)-f(a)}{z-a}.$$

A symmetric argument will then yield the existence of $f'_+(a)$ and the availability of the relation $f'_-(a) \leq f'_+(a)$. On the other hand, starting with $x < u \leq v < y$ in $\operatorname{int} I$, the same Theorem 1.3.1 yields

$$\frac{f(u)-f(x)}{u-x} \leq \frac{f(v)-f(x)}{v-x} \leq \frac{f(v)-f(y)}{v-y},$$

so letting $u \to x+$ and $v \to y-$, we obtain that $f'_+(x) \leq f'_-(y)$.

Because f admits finite lateral derivatives at each interior point, it will be continuous at each interior point. $\square$

By Theorem 1.3.3, every continuous convex function f (defined on a non-degenerate compact interval $[a,b]$) admits derivatives $f'_+(a)$ and $f'_-(b)$ at the endpoints, but they can be infinite,

$$-\infty \leq f'_+(a) < \infty \quad \text{and} \quad -\infty < f'_-(b) \leq \infty.$$

How nondifferentiable can a convex function be? Due to Theorem 1.3.3 above, we can immediately prove that every convex function $f\colon I \to \mathbb{R}$ is differentiable except for an enumerable subset. In fact, by considering the set

$$I_{\mathrm{nd}} = \{x \mid f'_-(x) < f'_+(x)\}$$

and letting for each $x \in I_{\mathrm{nd}}$ a rational point $r_x \in (f'_-(x), f'_+(x))$ we get a one-to-one function $\varphi\colon x \to r_x$ from I_{nd} into $\mathbb{Q}$. Consequently, I_{nd} is at most countable. Notice that this reasoning depends on the axiom of choice.

An example of a convex function which is not differentiable on a dense countable set will be exhibited in Remark 1.6.2 below. See also Exercise 3 at the end of this section.

Simple examples such as $f(x) = 0$ if $x \in (0,1)$, and $f(0) = f(1) = 1$, show that upward jumps could appear at the endpoints of the interval of definition of a convex function. Fortunately, the possible discontinuities are removable:

Proposition 1.3.4 *If $f\colon [a,b] \to \mathbb{R}$ is a convex function, then $f(a+)$ and $f(b-)$ exist in $\mathbb{R}$ and*

$$\widetilde{f}(x) = \begin{cases} f(a+) & \text{if } x = a \\ f(x) & \text{if } x \in (a,b) \\ f(b-) & \text{if } x = b \end{cases}$$

is convex too.

This result is a consequence of the following:

Proposition 1.3.5 *If $f\colon I \to \mathbb{R}$ is convex, then either f is monotonic on $\operatorname{int} I$, or there exists an $\xi \in \operatorname{int} I$ such that f is nonincreasing on the interval $(-\infty, \xi] \cap I$ and nondecreasing on the interval $[\xi, \infty) \cap I$.*

Proof. Since any convex function verifies formulas of the type (1.2), it suffices to consider the case where I is open. If f is not monotonic, then there must exist points $a < b < c$ in I such that

$$f(a) > f(b) < f(c).$$

The other possibility, $f(a) < f(b) > f(c)$, is rejected by the same formula (1.2). Since f is continuous on $[a,c]$, it attains its infimum on this interval at a point $\xi \in [a,c]$, that is,

$$f(\xi) = \inf f([a, c]).$$

Actually, $f(\xi) = \inf f(I)$. In fact, if $x < a$, then according to Theorem 1.3.1 we have

$$\frac{f(x) - f(\xi)}{x - \xi} \leq \frac{f(a) - f(\xi)}{a - \xi},$$

which yields $(\xi - a)f(x) \geq (x - a)f(\xi) + (\xi - x)f(a) \geq (\xi - a)f(\xi)$, that is, $f(x) \geq f(\xi)$. The other case, when $c < x$, can be treated in a similar manner.

If $u < v < \xi$, then

$$s_u(\xi) = s_\xi(u) \leq s_\xi(v) = \frac{f(v) - f(\xi)}{v - \xi} \leq 0$$

and thus $s_u(v) \leq s_u(\xi) \leq 0$. This shows that f is nonincreasing on $I \cap (-\infty, \xi]$. Analogously, if $\xi < u < v$, then from $s_v(\xi) \leq s_v(u)$ we infer that $f(v) \geq f(u)$, hence f is nondecreasing on $I \cap [\xi, \infty)$. $\square$

Corollary 1.3.6 *Every convex function $f \colon I \to \mathbb{R}$ which is not monotonic on* $\operatorname{int} I$ *has an interior global minimum.*

There is another way to look at the smoothness properties of the convex functions, based on the Lipschitz condition. A function f defined on an interval J is said to be *Lipschitz* if there exists a constant $L \geq 0$ such that

$$|f(x) - f(y)| \leq L|x - y| \quad \text{for all } x, y \in J.$$

A famous result due to H. Rademacher asserts that any Lipschitz function is differentiable almost everywhere. See Theorem 3.11.1.

Theorem 1.3.7 *If $f \colon I \to \mathbb{R}$ is a convex function, then f is Lipschitz on any compact interval $[a, b]$ contained in the interior of I.*

Proof. By Theorem 1.3.3,

$$f'_+(a) \leq f'_+(x) \leq \frac{f(y) - f(x)}{y - x} \leq f'_-(y) \leq f'_-(b)$$

for all $x, y \in [a, b]$ with $x < y$, hence $f|_{[a,b]}$ verifies the Lipschitz condition with $L = \max\{|f'_+(a)|, |f'_-(b)|\}$. $\square$

Corollary 1.3.8 *If $f_n \colon I \to \mathbb{R}$ $(n \in \mathbb{N})$ is a pointwise converging sequence of convex functions, then its limit f is also convex. Moreover, the convergence is uniform on any compact subinterval included in* $\operatorname{int} I$*, and $(f'_n)_n$ converges to f' except possibly at countably many points of I.*

Since the first derivative of a convex function may not exist at a dense subset, a characterization of convexity in terms of second order derivatives is not possible unless we relax the concept of twice differentiability. The *upper* and the *lower second symmetric derivative* of f at x, are respectively defined by the formulas

$$\overline{\mathcal{D}}^2 f(x) = \limsup_{h\downarrow 0} \frac{f(x+h)+f(x-h)-2f(x)}{h^2}$$
$$\underline{\mathcal{D}}^2 f(x) = \liminf_{h\downarrow 0} \frac{f(x+h)+f(x-h)-2f(x)}{h^2}.$$

It is not difficult to check that if f is twice differentiable at a point x, then

$$\overline{\mathcal{D}}^2 f(x) = \underline{\mathcal{D}}^2 f(x) = f''(x);$$

however $\overline{\mathcal{D}}^2 f(x)$ and $\underline{\mathcal{D}}^2 f(x)$ can exist even at points of discontinuity; for example, consider the case of the signum function and the point $x = 0$.

Theorem 1.3.9 *Suppose that I is an open interval. A real-valued function f is convex on I if and only if f is continuous and $\overline{\mathcal{D}}^2 f \geq 0$.*

Accordingly, if a function $f\colon I \to \mathbb{R}$ is convex in the neighborhood of each point of I, then it is convex on the whole interval I.

Proof. If f is convex, then clearly $\overline{\mathcal{D}}^2 f \geq \underline{\mathcal{D}}^2 f \geq 0$. The continuity of f follows from Theorem 1.3.3.

Now, suppose that $\overline{\mathcal{D}}^2 f > 0$ on I. If f is not convex, then we can find a point x_0 such that $\overline{\mathcal{D}}^2 f(x_0) \leq 0$, which will be a contradiction. In fact, in this case there exists a subinterval $I_0 = [a_0, b_0]$ such that $f((a_0+b_0)/2) > (f(a_0)+f(b_0))/2$. A moment's reflection shows that one of the intervals $[a_0, (a_0+b_0)/2]$, $[(3a_0+b_0)/4, (a_0+3b_0)/4]$, $[(a_0+b_0)/2, b_0]$ can be chosen to replace I_0 by a smaller interval $I_1 = [a_1, b_1]$, with $b_1 - a_1 = (b_0 - a_0)/2$ and $f((a_1+b_1)/2) > (f(a_1)+f(b_1))/2$. Proceeding by induction, we arrive at a situation where the principle of included intervals gives us the point x_0.

In the general case, consider the sequence of functions

$$f_n(x) = f(x) + \frac{1}{n}x^2.$$

Then $\overline{\mathcal{D}}^2 f_n > 0$, and the above reasoning shows us that f_n is convex. Clearly $f_n(x) \to f(x)$ for each $x \in I$, so that the convexity of f will be a consequence of Corollary 1.3.8 above. $\square$

Corollary 1.3.10 (The second derivative test) *Suppose that $f\colon I \to \mathbb{R}$ is a twice differentiable function. Then:*

(i) *f is convex if and only if $f'' \geq 0$;*

(ii) *f is strictly convex if and only if $f'' \geq 0$ and the set of points where f'' vanishes does not include intervals of positive length.*

An important result due to A. D. Alexandrov asserts that all convex functions are almost everywhere twice differentiable. See Theorem 3.11.2.

Remark 1.3.11 (Higher order convexity) The following generalization of the notion of a convex function was initiated by T. Popoviciu in 1934. A function $f\colon [a,b] \to \mathbb{R}$ is said to be *n-convex* ($n \in \mathbb{N}^*$) if for all choices of $n+1$ distinct points $x_0 < \cdots < x_n$ in $[a,b]$, the n-th order divided difference of f satisfies

$$f[x_0, \ldots, x_n] \geq 0.$$

The divided differences are given inductively by

$$\begin{aligned}
f[x_0, x_1] &= \frac{f(x_0) - f(x_1)}{x_0 - x_1} \\
f[x_0, x_1, x_2] &= \frac{f[x_0, x_1] - f[x_1, x_2]}{x_0 - x_2} \\
&\vdots \\
f[x_0, \ldots, x_n] &= \frac{f[x_0, \ldots, x_{n-1}] - f[x_1, \ldots, x_n]}{x_0 - x_n}.
\end{aligned}$$

Thus the 1-convex functions are the nondecreasing functions, while the 2-convex functions are precisely the classical convex functions. In fact,

$$\begin{vmatrix} 1 & x & f(x) \\ 1 & y & f(y) \\ 1 & z & f(z) \end{vmatrix} \Bigg/ \begin{vmatrix} 1 & x & x^2 \\ 1 & y & y^2 \\ 1 & z & z^2 \end{vmatrix} = \frac{f[y,z] - f[x,z]}{y - x},$$

and the claim follows from Lemma 1.3.2. As T. Popoviciu noticed in his book [205], if f is n-times differentiable, with $f^{(n)} \geq 0$, then f is n-convex.

See [196] and [212] for a more detailed account on the theory of n-convex functions.

Exercises

1. (An application of the second derivative test of convexity)

(i) Prove that the functions $\log((e^{ax}-1)/(e^x-1))$ and $\log(\sinh ax/\sinh x)$ are convex on $\mathbb{R}$ if $a \geq 1$.

(ii) Prove that the function $b \log\cos(x/\sqrt{b}) - a\log\cos(x/\sqrt{a})$ is convex on $(0, \pi/2)$ if $b \geq a \geq 1$.

2. Suppose that $0 < a < b < c$ (or $0 < b < c < a$, or $0 < c < b < a$). Use Lemma 1.3.2 to infer the following inequalities:

(i) $ab^\alpha + bc^\alpha + ca^\alpha > ac^\alpha + ba^\alpha + cb^\alpha$ for $\alpha \geq 1$;
(ii) $a^b b^c c^a > a^c c^b b^a$;
(iii) $$\frac{a(c-b)}{(c+b)(2a+b+c)} + \frac{b(a-c)}{(a+c)(a+2b+c)} + \frac{c(b-a)}{(b+a)(a+b+2c)} > 0.$$

3. Show that the function $f(x) = \sum_{n=0}^{\infty} |x-n|/2^n$, $x \in \mathbb{R}$, provides an example of a convex function which is nondifferentiable on a countable subset.

4. Let D be a bounded closed convex subset of the real plane. Prove that D can be always represented as

$$D = \{(x,y) \mid f(x) \leq y \leq g(x),\, x \in [a,b]\}$$

for suitable functions $f\colon [a,b] \to \mathbb{R}$ convex, and $g\colon [a,b] \to \mathbb{R}$ concave. Infer that the boundary of D is smooth except possibly for an enumerable subset.

5. Prove that a continuous convex function $f\colon [a,b] \to \mathbb{R}$ can be extended to a convex function on $\mathbb{R}$ if and only if $f'_+(a)$ and $f'_-(b)$ are finite.

6. Use Corollary 1.3.10 to prove that the sine function is strictly concave on $[0,\pi]$. Infer that

$$\left(\frac{\sin a}{a}\right) x \leq \sin x \leq \sin a \left(\frac{x}{a}\right)^{a \cot a}$$

for every $a \in (0, \pi/2]$ and every $x \in [0,a]$. For $a = \pi/2$ this yields the classical *inequality of Jordan.*

7. Let $f\colon [0,2\pi] \to \mathbb{R}$ be a convex function. Prove that

$$a_n = \frac{1}{\pi} \int_0^{2\pi} f(t) \cos nt \, dt \geq 0 \quad \text{for every } n \geq 1.$$

8. (J. L. W. V. Jensen [115]) Prove that a function $f\colon [0,M] \to \mathbb{R}$ is nondecreasing if and only if

$$\sum_{k=1}^{n} \lambda_k f(x_k) \leq \Big(\sum_{k=1}^{n} \lambda_k\Big) f\Big(\sum_{k=1}^{n} x_k\Big)$$

for all finite families $\lambda_1, \ldots, \lambda_n \geq 0$ and $x_1, \ldots, x_n \in [0,M]$, with $\sum_{k=1}^n x_k \leq M$ and $n \geq 2$. This applies to any continuous convex function $g\colon [0,M] \to \mathbb{R}$, noticing that $[g(x) - g(0)]/x$ is nondecreasing.

9. (van der Corput's lemma) Let $\lambda > 0$ and let $f\colon \mathbb{R} \to \mathbb{R}$ be a function of class C^2 such that $f'' \geq \lambda$. Prove that

$$\left| \int_a^b \mathrm{e}^{\mathrm{i}f(t)}\, dt \right| \leq 4\sqrt{2/\lambda} \quad \text{for all } a, b \in \mathbb{R}.$$

[*Hint*: Use integration by parts on intervals around the point where f' vanishes.]

1.4 An Upper Estimate of Jensen's Inequality

An important topic related to inequalities is their precision. The following result (which exhibits the power of one-variable techniques in a several-variables context) yields an upper estimate of Jensen's inequality:

Theorem 1.4.1 *Let* $f\colon [a,b] \to \mathbb{R}$ *be a convex function and let*

$$[m_1, M_1], \ldots, [m_n, M_n]$$

be compact subintervals of $[a,b]$*. Given* $\lambda_1, \ldots, \lambda_n$ *in* $[0,1]$*, with* $\sum_{k=1}^n \lambda_k = 1$*, the function*

$$E(x_1, \ldots, x_n) = \sum_{k=1}^n \lambda_k f(x_k) - f\Big(\sum_{k=1}^n \lambda_k x_k\Big)$$

attains its maximum on $\Omega = [m_1, M_1] \times \cdots \times [m_n, M_n]$ *at a vertex, that is, at a point of* $\{m_1, M_1\} \times \cdots \times \{m_n, M_n\}$*.*

The proof depends upon the following refinement of Lagrange's mean value theorem:

Lemma 1.4.2 *Let* $h\colon [a,b] \to \mathbb{R}$ *be a continuous function. Then there exists a point* $c \in (a,b)$ *such that*

$$\underline{D}h(c) \le \frac{h(b) - h(a)}{b - a} \le \overline{D}h(c).$$

Here

$$\underline{D}h(c) = \liminf_{x \to c} \frac{h(x) - h(c)}{x - c} \quad \text{and} \quad \overline{D}h(c) = \limsup_{x \to c} \frac{h(x) - h(c)}{x - c}.$$

are respectively the lower derivative *and the* upper derivative *of* h *at* c*.*

Proof. As in the smooth case, we consider the function

$$H(x) = h(x) - \frac{h(b) - h(a)}{b - a}\,(x - a), \quad x \in [a,b].$$

Clearly, H is continuous and $H(a) = H(b)$. If H attains its supremum at $c \in (a,b)$, then $\underline{D}H(c) \le 0 \le \overline{D}H(c)$ and the conclusion of Lemma 1.4.2 is immediate. The same is true when H attains its infimum at an interior point of $[a,b]$. If both extremes are attained at the endpoints, then H is constant and the conclusion of Lemma 1.4.2 works for all c in (a,b). $\square$

Proof of Theorem 1.4.1. Clearly, we may assume that f is also continuous. We shall show (by *reductio ad absurdum*) that

$$E(x_1,\ldots,x_k,\ldots,x_n)\le\sup\{E(x_1,\ldots,m_k,\ldots,x_n),E(x_1,\ldots,M_k,\ldots,x_n)\}$$

for all $(x_1,x_2,\ldots,x_n)\in\Omega$ and all $k\in\{1,\ldots,n\}$. In fact, if

$$E(x_1,x_2,\ldots,x_n)>\sup\{E(m_1,x_2,\ldots,x_n),E(M_1,x_2,\ldots,x_n)\}$$

for some $(x_1,x_2,\ldots,x_n)\in\Omega$, we consider the function

$$h\colon[m_1,M_1]\to\mathbb{R},\quad h(x)=E(x,x_2,\ldots,x_n).$$

According to Lemma 1.4.2, there exists a $\xi\in(m_1,x_1)$ such that

$$h(x_1)-h(m_1)\le(x_1-m_1)\overline{D}h(\xi).$$

Since $h(x_1)>h(m_1)$, it follows that $\overline{D}h(\xi)>0$, equivalently,

$$\overline{D}f(\xi)>\overline{D}f(\lambda_1\xi+\lambda_2x_2+\cdots+\lambda_nx_n).$$

Or, $\overline{D}f=f'_+$ is a nondecreasing function on (a,b), which yields

$$\xi>\lambda_1\xi+\lambda_2x_2+\cdots+\lambda_nx_n,$$

and thus $\xi>(\lambda_2x_2+\cdots+\lambda_nx_n)/(\lambda_2+\cdots+\lambda_n)$.

A new appeal to Lemma 1.4.2 (applied this time to $h|_{[x_1,M_1]}$), yields an $\eta\in(x_1,M_1)$ such that $\eta<(\lambda_2x_2+\cdots+\lambda_nx_n)/(\lambda_2+\cdots+\lambda_n)$. But this contradicts the fact that $\xi<\eta$. □

Corollary 1.4.3 *Let $f\colon[a,b]\to\mathbb{R}$ be a convex function. Then*

$$\frac{f(a)+f(b)}{2}-f\Big(\frac{a+b}{2}\Big)\ge\frac{f(c)+f(d)}{2}-f\Big(\frac{c+d}{2}\Big)$$

for all $a\le c\le d\le b$.

An application of Corollary 1.4.3 to series summation may be found in [99, p. 100].

Theorem 1.4.1 allows us to retrieve a remark due to L. G. Khanin [125]: *Let $p>1$, $x_1,\ldots,x_n\in[0,M]$ and $\lambda_1,\ldots,\lambda_n\in[0,1]$, with $\sum_{k=1}^n\lambda_k=1$. Then*

$$\sum_{k=1}^n\lambda_kx_k^p\le\Big(\sum_{k=1}^n\lambda_kx_k\Big)^p+(p-1)p^{p/(1-p)}M^p.$$

In particular,

$$\frac{x_1^2+\cdots+x_n^2}{n}\le\Big(\frac{x_1+\cdots+x_n}{n}\Big)^2+\frac{M^2}{4},$$

which represents an additive converse to the Cauchy–Buniakovski–Schwarz inequality. In fact, according to Theorem 1.4.1, the function

$$E(x_1,\ldots,x_n)=\sum_{k=1}^n\lambda_kx_k^p-\Big(\sum_{k=1}^n\lambda_kx_k\Big)^p,$$

attains its supremum on $[0,M]^n$ at a point whose coordinates are either 0 or M. Therefore $\sup E(x_1,\ldots,x_n)$ does not exceed $M^p\cdot\sup\{s-s^p\mid s\in[0,1]\}=(p-1)\,p^{p/(1-p)}M^p$.

Exercises

1. (Kantorovich's inequality) Let $m, M, a_1, \ldots, a_n$ be positive numbers, with $m < M$. Prove that the maximum of

$$f(x_1, \ldots, x_n) = \Big(\sum_{k=1}^{n} a_k x_k\Big)\Big(\sum_{k=1}^{n} a_k/x_k\Big)$$

for $x_1, \ldots, x_n \in [m, M]$ is equal to

$$\frac{(M+m)^2}{4Mm}\Big(\sum_{k=1}^{n} a_k\Big)^2 - \frac{(M-m)^2}{4Mm}\min_{X \subset \{1,\ldots,n\}}\Big(\sum_{k\in X} a_k - \sum_{k \in \complement X} a_k\Big)^2.$$

Remark. The following particular case

$$\Big(\frac{1}{n}\sum_{k=1}^{n} x_k\Big)\Big(\frac{1}{n}\sum_{k=1}^{n}\frac{1}{x_k}\Big) \leq \frac{(M+m)^2}{4Mm} - \frac{(1+(-1)^{n+1})(M-m)^2}{8Mmn^2}$$

represents an improvement on Schweitzer's inequality for odd n.

2. Let $a_k, b_k, c_k, m_k, M_k, m'_k, M'_k$ be positive numbers with $m_k < M_k$ and $m'_k < M'_k$ for $k \in \{1, \ldots, n\}$ and let $p > 1$. Prove that the maximum of

$$\Big(\sum_{k=1}^{n} a_k x_k^p\Big)\Big(\sum_{k=1}^{n} b_k y_k^p\Big)\Big/\Big(\sum_{k=1}^{n} c_k x_k y_k\Big)^p$$

for $x_k \in [m_k, M_k]$ and $y_k \in [m'_k, M'_k]$ ($k \in \{1, \ldots, n\}$) is attained at a $2n$-tuple whose components are endpoints.

3. Assume that $f\colon I \to \mathbb{R}$ is strictly convex and continuous and $g\colon I \to \mathbb{R}$ is continuous. For $a_1, \ldots, a_n > 0$ and $m_k, M_k \in I$, with $m_k < M_k$ for $k \in \{1, \ldots, n\}$, consider the function

$$h(x_1, \ldots, x_n) = \sum_{k=1}^{n} a_k f(x_k) + g\Big(\sum_{k=1}^{n} a_k x_k\Big/\sum_{k=1}^{n} a_k\Big)$$

defined on $\prod_{k=1}^{n}[m_k, M_k]$. Prove that a necessary condition for a point $(y_1, \ldots, y_n)$ to be a point of maximum is that at most one component y_k is inside the corresponding interval $[m_k, M_k]$.

1.5 The Subdifferential

In the case of nonsmooth convex functions, the lack of tangent lines can be supplied by support lines. See Fig. 1.3. Given a function $f\colon I \to \mathbb{R}$, we say that f admits a *support line* at $x \in I$ if there exists a $\lambda \in \mathbb{R}$ such that

$$f(y) \geq f(x) + \lambda(y - x), \quad \text{for all } y \in I.$$

We call the set $\partial f(x)$ of all such λ the *subdifferential of f at x*. Geometrically, the subdifferential gives us the slopes of the supporting lines for the graph of f. The subdifferential is always a convex set, possibly empty.

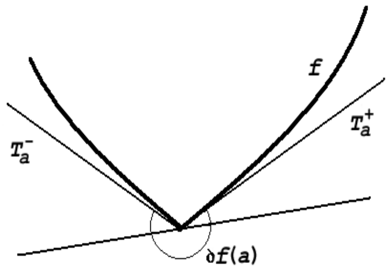

Fig. 1.3. Convexity: the existence of support lines at interior points.

The convex functions have the remarkable property that $\partial f(x) \neq \emptyset$ at all interior points. However, even in their case, the subdifferential could be empty at the endpoints. An example is given by the continuous convex function $f(x) = 1 - \sqrt{1 - x^2}$, $x \in [-1, 1]$, which fails to have a support line at $x = \pm 1$.

We may think of $\partial f(x)$ as the value at x of a set-valued function ∂f (the *subdifferential* of f), whose domain $\operatorname{dom} \partial f$ consists of all points x in I where f has a support line.

Lemma 1.5.1 *Let f be a convex function on an interval I. Then $\partial f(x) \neq \emptyset$ at all interior points of I. Moreover, every function $\varphi\colon I \to \mathbb{R}$ for which $\varphi(x) \in \partial f(x)$ whenever $x \in \operatorname{int} I$ verifies the double inequality*

$$f'_-(x) \leq \varphi(x) \leq f'_+(x),$$

and thus is nondecreasing on $\operatorname{int} I$.

The conclusion above includes the endpoints of I provided that f is differentiable there. As a consequence, the differentiability of a convex function f at a point means that f admits a unique support line at that point.

Proof. First, we shall prove that $f'_+(a) \in \partial f(a)$ for each $a \in \operatorname{int} I$ (and also at the leftmost point of I, provided that f is differentiable there). In fact, if $x \in I$, with $x \geq a$, then

$$\frac{f((1-t)a + tx) - f(a)}{t} \leq f(x) - f(a)$$

for all $t \in (0, 1]$, which yields

$$f(x) \geq f(a) + f'_+(a) \cdot (x - a).$$

If $x \leq a$, then a similar argument leads us to $f(x) \geq f(a) + f'_-(a) \cdot (x - a)$; or $f'_-(a) \cdot (x - a) \geq f'_+(a) \cdot (x - a)$, because $x - a \leq 0$.

Analogously, we can argue that $f'_-(a) \in \partial f(a)$ for all $a \in \operatorname{int} I$ (and also for the rightmost point in I provided that f is differentiable at that point).

The fact that φ is nondecreasing follows now from Theorem 1.3.3. $\square$

Every continuous convex function is the upper envelope of its support lines. More precisely:

Theorem 1.5.2 *Let f be a continuous convex function on an interval I and let $\varphi\colon I \to \mathbb{R}$ be a function such that $\varphi(x)$ belongs to $\partial f(x)$ for all $x \in \operatorname{int} I$. Then*

$$f(z) = \sup\{f(x) + (z - x)\varphi(x) \mid x \in \operatorname{int} I\} \quad \textit{for all } z \in I.$$

Proof. The case of interior points is clear. If z is an endpoint, say the left one, then we have already noticed that

$$f(z + t) - f(z) \leq t\varphi(z + t) \leq f(z + 2t) - f(z + t)$$

for $t > 0$ small enough, which yields $\lim_{t \to 0+} t\varphi(z + t) = 0$. Given $\varepsilon > 0$, there is $\delta > 0$ such that $|f(z) - f(z + t)| < \varepsilon/2$ and $|t\varphi(z + t)| < \varepsilon/2$ for $0 < t < \delta$. This shows that $f(z + t) - t\varphi(z + t) < f(z) + \varepsilon$ for $0 < t < \delta$ and the result follows. $\square$

The following result shows that only the convex functions satisfy the condition $\partial f(x) \neq \emptyset$ at all interior points of I.

Theorem 1.5.3 *Let $f\colon I \to \mathbb{R}$ be a function such that $\partial f(x) \neq \emptyset$ at all interior points x of I. Then f is convex.*

Proof. Let $u, v \in I$, $u \neq v$, and let $t \in (0, 1)$. Then $(1 - t)u + tv \in \operatorname{int} I$, so that for all $\lambda \in \partial f((1 - t)u + tv)$ we get

$$\begin{aligned} f(u) &\geq f((1 - t)u + tv) + t(u - v) \cdot \lambda, \\ f(v) &\geq f((1 - t)u + tv) - (1 - t)(u - v) \cdot \lambda. \end{aligned}$$

By multiplying the first inequality by $1 - t$, the second one by t and then adding them side by side, we get $(1 - t)f(u) + tf(v) \geq f((1 - t)u + tv)$, hence f is a convex function. $\square$

We shall illustrate the importance of the subdifferential by proving two classical results. The first one is the basis of the theory of majorization, which will be later described in Section 1.10.

Theorem 1.5.4 (The Hardy–Littlewood–Pólya inequality) *Suppose that f is a convex function on an interval I and consider two families $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ of points in I such that*

$$\sum_{k=1}^{m} x_k \le \sum_{k=1}^{m} y_k \quad \text{for } m \in \{1, \ldots, n\}$$

and

$$\sum_{k=1}^{n} x_k = \sum_{k=1}^{n} y_k.$$

If $x_1 \ge \cdots \ge x_n$, then

$$\sum_{k=1}^{n} f(x_k) \le \sum_{k=1}^{n} f(y_k),$$

while if $y_1 \le \cdots \le y_n$ this inequality works in the reverse direction.

Proof. We shall concentrate here on the first conclusion (concerning the decreasing families), which will be settled by mathematical induction. The second conclusion follows from the first one by replacing f by $\tilde{f} \colon \tilde{I} \to \mathbb{R}$, where $\tilde{I} = \{-x \mid x \in I\}$ and $\tilde{f}(x) = f(-x)$ for $x \in \tilde{I}$.

The case $n = 1$ is clear. Assuming the conclusion valid for all families of length $n-1$, we pass to the case of families of length n. Under the hypotheses of Theorem 1.5.4, we have $x_1, x_2, \ldots, x_n \in [\min_k y_k, \max_k y_k]$, so that we may restrict to the case where

$$\min_k y_k < x_1, \ldots, x_n < \max_k y_k.$$

Then $x_1, \ldots, x_n$ are interior points of I. According to Lemma 1.5.1 we may choose a nondecreasing function $\varphi \colon \operatorname{int} I \to \mathbb{R}$ such that $\varphi(x) \in \partial f(x)$ for all $x \in \operatorname{int} I$. By Theorem 1.5.2 and Abel's summation formula we get

$$\begin{aligned}
&\sum_{k=1}^{n} f(y_k) - \sum_{k=1}^{n} f(x_k) \\
&\qquad \ge \sum_{k=1}^{n} \varphi(x_k)(y_k - x_k) \\
&\qquad = \varphi(x_1)(y_1 - x_1) + \sum_{m=2}^{n} \varphi(x_m)\Big[\sum_{k=1}^{m}(y_k - x_k) - \sum_{k=1}^{m-1}(y_k - x_k)\Big] \\
&\qquad = \varphi(x_n)\sum_{k=1}^{n}(y_k - x_k) + \sum_{m=1}^{n-1}\Big[(\varphi(x_m) - \varphi(x_{m+1}))\sum_{k=1}^{m}(y_k - x_k)\Big] \\
&\qquad = \sum_{m=1}^{n-1}\Big[(\varphi(x_m) - \varphi(x_{m+1}))\sum_{k=1}^{m}(y_k - x_k)\Big]
\end{aligned}$$

and the proof is complete. □

A more general result (which includes also Corollary 1.4.3) will be the objective of Theorem 2.7.8.

Remark 1.5.5 The Hardy–Littlewood–Pólya inequality implies many other inequalities on convex functions. We shall detail here the case of Popoviciu's inequality (see Theorem 1.1.8).

Without loss of generality we may assume the ordering $x \geq y \geq z$. Then

$$(x+y)/2 \geq (z+x)/2 \geq (y+z)/2 \quad \text{and} \quad x \geq (x+y+z)/3 \geq z.$$

If $x \geq (x+y+z)/3 \geq y \geq z$, then the conclusion of Theorem 1.1.8 follows from Theorem 1.5.4, applied to the families

$$\begin{gathered} x_1 = x, \quad x_2 = x_3 = x_4 = (x+y+z)/3, \quad x_5 = y, \quad x_6 = z, \\ y_1 = y_2 = (x+y)/2, \quad y_3 = y_4 = (x+z)/2, \quad y_5 = y_6 = (y+z)/2, \end{gathered}$$

while in the case $x \geq y \geq (x+y+z)/3 \geq z$, we have to consider the families

$$\begin{gathered} x_1 = x, \quad x_2 = y, \quad x_3 = x_4 = x_5 = (x+y+z)/3, \quad x_6 = z, \\ y_1 = y_2 = (x+y)/2, \quad y_3 = y_4 = (x+z)/2, \quad y_5 = y_6 = (y+z)/2. \end{gathered}$$

Our second application concerns a classical generalization of Jensen's inequality, which deals with linear (not necessarily convex) combinations:

Theorem 1.5.6 (The Jensen–Steffensen inequality) *Let $x_n \leq x_{n-1} \leq \cdots \leq x_1$ be points in $[a,b]$ and let $p_1, \ldots, p_n$ be real numbers such that the partial sums $S_k = \sum_{i=1}^{k} p_i$ verify the relations*

$$0 \leq S_k \leq S_n \quad \textit{and} \quad S_n > 0.$$

Then every convex function f defined on $[a,b]$ verifies the inequality

$$f\Big(\frac{1}{S_n}\sum_{k=1}^{n} p_k x_k\Big) \leq \frac{1}{S_n}\sum_{k=1}^{n} p_k f(x_k).$$

Proof. Put $\bar{x} = (\sum_{k=1}^{n} p_k x_k)/S_n$ and let $\bar{S}_k = S_n - S_{k-1} = \sum_{i=k}^{n} p_i$. Then

$$S_n(x_1 - \bar{x}) = \sum_{i=1}^{n} p_i(x_1 - x_i) = \sum_{j=2}^{n} (x_{j-1} - x_j)\bar{S}_j \geq 0$$

and

$$S_n(\bar{x} - x_n) = \sum_{i=1}^{n-1} p_i(x_i - x_n) = \sum_{j=1}^{n-1} (x_j - x_{j+1})S_j \geq 0,$$

which shows that $x_n \leq \bar{x} \leq x_1$. At this point we may restrict ourselves to the case where f is continuous and the points $x_1, \ldots, x_n$ belong to (a, b). See Proposition 1.3.4. According to Lemma 1.5.1, we may choose a function $\varphi\colon I \to \mathbb{R}$ such that $\varphi(x) \in \partial f(x)$ for all $x \in \operatorname{int} I$. Then

$$f(z) - f(y) \geq \varphi(c)(z - y) \quad \text{if } z \geq y \geq c$$

and

$$f(z) - f(y) \leq \varphi(c)(z - y) \quad \text{if } c \geq z \geq y.$$

Choose also an index m such that $\bar{x} \in [x_{m+1}, x_m]$. Then

$$f\Big(\frac{1}{S_n}\sum_{k=1}^{n} p_k x_k\Big) - \frac{1}{S_n}\sum_{k=1}^{n} p_k f(x_k)$$

is majorized by

$$\begin{aligned}
&\sum_{i=1}^{m-1} [\varphi(\bar{x})(x_i - x_{i+1}) - f(x_i) + f(x_{i+1})] \frac{S_i}{S_n} \\
&\quad + [\varphi(\bar{x})(x_m - \bar{x}) - f(x_m) + f(\bar{x})] \frac{S_m}{S_n} \\
&\quad + [f(\bar{x}) - f(x_{m+1}) - \varphi(\bar{x})(\bar{x} - x_{m+1})] \frac{\bar{S}_{m+1}}{S_n} \\
&\quad + \sum_{i=m+1}^{n-1} [f(x_i) - f(x_{i+1}) - \varphi(\bar{x})(x_i - x_{i+1})] \frac{\bar{S}_{i+1}}{S_n},
\end{aligned}$$

which is a sum of nonpositive numbers. $\square$

A powerful device to prove inequalities for convex functions is to take advantage of some structure results. The class of piecewise linear convex functions appears to be very important. Here a function $f\colon [a, b] \to \mathbb{R}$ is said to be *piecewise linear* if there exists a division $a = x_0 < \cdots < x_n = b$ such that the restriction of f to each partial interval $[x_k, x_{k+1}]$ is an affine function.

Theorem 1.5.7 (T. Popoviciu [203]) *Let $f\colon [a, b] \to \mathbb{R}$ be a piecewise linear convex function. Then f is the sum of an affine function and a linear combination, with positive coefficients, of translates of the absolute value function. In other words, f is of the form*

$$f(x) = \alpha x + \beta + \sum_{k=1}^{n} c_k |x - x_k| \tag{1.11}$$

for suitable $\alpha, \beta \in \mathbb{R}$ and suitable nonnegative coefficients $c_1, \ldots, c_n$.

Proof. Let $a = x_0 < \cdots < x_m = b$ be a division of $[a, b]$ such that the restriction of f to each partial interval $[x_k, x_{k+1}]$ is affine. If $\alpha x + \beta$ is the affine function whose restriction to $[x_0, x_1]$ coincides with $f|_{[x_0,x_1]}$, then it will be a support line for f and $f(x) - (\alpha x + \beta)$ will be a nondecreasing convex function which vanishes on $[x_0, x_1]$. A moment's reflection shows the existence of a constant $c_1 \geq 0$ such that $f(x) - (\alpha x + \beta) = c_1(x - x_1)^+$ on $[x_0, x_2]$. Repeating the argument we arrive at the representation

$$f(x) = \alpha x + \beta + \sum_{k=1}^{m-1} c_k(x - x_k)^+ \tag{1.12}$$

where all coefficients c_k are nonnegative. The proof ends by replacing the translates of the positive part function by translates of the absolute value function. This is possible via the formula $y^+ = (|y| + y)/2$. $\square$

Suppose that we want to prove the validity of the discrete form of Jensen's inequality for all continuous convex functions $f\colon [a, b] \to \mathbb{R}$. Since every such function can be uniformly approximated by piecewise linear convex functions we may restrict ourselves to this particular class of functions. If Jensen's inequality works for two functions f_1 and f_2, it also works for every combination $c_1 f_1 + c_2 f_2$ with nonnegative coefficients. According to Theorem 1.5.7, this shows that the proof of Jensen's inequality (within the class of continuous convex functions $f\colon [a, b] \to \mathbb{R}$) reduces to its verification for affine functions and translates $x \to |x - y|$, or both cases are immediate. In the same manner (but using the representation formula (1.12)), one can prove the Hardy–Littlewood–Pólya inequality. Popoviciu's inequality was originally proved via Theorem 1.5.7. For it, the case of the absolute value function reduces to Hlawka's inequality on the real line, that is,

$$|x| + |y| + |z| + |x + y + z| \geq |x + y| + |y + z| + |z + x|$$

for all $x, y, z \in \mathbb{R}$. A simple proof may be found in the Comments section at the end of Chapter 2.

The representation formula (1.11) admits a generalization for all continuous convex functions on intervals. See Theorem 1.6.3.

Exercises

1. Let $f\colon I \to \mathbb{R}$ be a convex function. Show that:

(i) Any local minimum of f is a global one;
(ii) f attains a global minimum at a if and only if $0 \in \partial f(a)$;
(iii) if f has a global maximum at an interior point of I, then f is constant.

2. Suppose that $f\colon \mathbb{R} \to \mathbb{R}$ is a convex function which is bounded from above. Prove that f is constant.

3. (Convex mean value theorem) Consider a continuous convex function $f\colon [a,b] \to \mathbb{R}$. Prove that $(f(b)-f(a))/(b-a) \in \partial f(c)$ for some point $c \in (a,b)$.

4. (A geometric application of the Hardy–Littlewood–Pólya inequality; see M. S. Klamkin [127]) Let P, A and P', A' denote the perimeter and area, respectively, of two convex polygons $\mathcal{P}$ and $\mathcal{P}'$ inscribed in the same circle (the center of the circle lies in the interior of both polygons). If the greatest side of $\mathcal{P}'$ is less than or equal with the smallest side of $\mathcal{P}$, prove that

$$P' \geq P \quad \text{and} \quad A' \geq A$$

with equality if and only if the polygons are congruent and regular.
[*Hint*: Express the perimeter and area of a polygon via the central angles subtended by the sides. Then use Theorem 1.5.4.]

5. (A. F. Berezin) Let P be an orthogonal projection in $\mathbb{R}^n$ and let A be a self-adjoint linear operator in $\mathbb{R}^n$. Infer from Theorem 1.5.7 that

$$\operatorname{trace}(Pf(PAP)P) \leq \operatorname{trace}(Pf(A)P)$$

for every convex function $f\colon \mathbb{R} \to \mathbb{R}$.

1.6 Integral Representation of Convex Functions

It is well known that differentiation and integration are operations inverse to each other. A consequence of this fact is the existence of a certain duality between the class of convex functions on an open interval and the class of nondecreasing functions on that interval.

Given a nondecreasing function $\varphi\colon I \to \mathbb{R}$ and a point $c \in I$ we can attach to them a new function f, given by

$$f(x) = \int_c^x \varphi(t)\,dt.$$

As φ is bounded on bounded intervals, it follows that f is locally Lipschitz (and thus continuous). It is also a convex function. In fact, according to Theorem 1.1.4, it suffices to show that f is midpoint convex. Or, for $x \leq y$ in I we have

$$\frac{f(x)+f(y)}{2} - f\Big(\frac{x+y}{2}\Big) = \frac{1}{2}\bigg(\int_{(x+y)/2}^{y} \varphi(t)\,dt - \int_{x}^{(x+y)/2} \varphi(t)\,dt\bigg) \geq 0$$

since φ is nondecreasing.

It is elementary that f is differentiable at each point of continuity of φ and $f' = \varphi$ at such points.

On the other hand, the subdifferential allows us to state the following generalization of the fundamental formula of integral calculus:

Proposition 1.6.1 *Let $f\colon I \to \mathbb{R}$ be a continuous convex function and let $\varphi\colon I \to \mathbb{R}$ be a function such that $\varphi(x) \in \partial f(x)$ for all $x \in \operatorname{int} I$. Then for all $a < b$ in I we have*

$$f(b) - f(a) = \int_a^b \varphi(t)\,dt.$$

Proof. Clearly, we may restrict ourselves to the case where $[a, b] \subset \operatorname{int} I$. If $a = t_0 < t_1 < \cdots < t_n = b$ is a division of $[a, b]$, then

$$f'_-(t_{k-1}) \le f'_+(t_{k-1}) \le \frac{f(t_k) - f(t_{k-1})}{t_k - t_{k-1}} \le f'_-(t_k) \le f'_+(t_k)$$

for all k. Since

$$f(b) - f(a) = \sum_{k=1}^{n} [f(t_k) - f(t_{k-1})],$$

a moment's reflection shows that

$$f(b) - f(a) = \int_a^b f'_-(t)\,dt = \int_a^b f'_+(t)\,dt.$$

On the other hand $f'_- \le \varphi \le f'_+$, which forces the equality in the statement of Proposition 1.6.1. ☐

Remark 1.6.2 There exist convex functions whose first derivative fails to exist on a dense set. For this, let $r_1, r_2, r_3, \ldots$ be an enumeration of the rational numbers in $[0, 1]$ and put

$$\varphi(t) = \sum_{\{k \mid r_k \le t\}} \frac{1}{2^k}.$$

Then

$$f(x) = \int_0^x \varphi(t)\,dt$$

is a continuous convex function whose first derivative does not exist at the points r_k. F. Riesz exhibited an example of increasing function φ with $\varphi' = 0$ almost everywhere. See [103, pp. 278–282]. The corresponding function f in his example is strictly convex though $f'' = 0$ almost everywhere. As we shall see in the Comments at the end of this chapter, Riesz's example is typical from the generic point of view.

We shall derive from Proposition 1.6.1 an important integral representation of all continuous convex functions $f\colon [a, b] \to \mathbb{R}$. For this we need the following Green function associated to the bounded open interval (a, b):

$$G(x, y) = \begin{cases} (x-a)(y-b)/(b-a), & \text{if } a \le x \le y \le b \\ (x-b)(y-a)/(b-a), & \text{if } a \le y \le x \le b. \end{cases}$$

Notice that G is continuous, symmetric and $G \leq 0$ on $[a,b] \times [a,b]$. It is a convex function in each variable (when the other is fixed) and vanishes at the boundary. Moreover,

$$\frac{\partial G}{\partial x}(x+0,x) - \frac{\partial G}{\partial x}(x-0,x) = 1.$$

Theorem 1.6.3 *For every continuous convex function $f\colon [a,b] \to \mathbb{R}$ there exists a uniquely determined positive Borel measure μ on $I = (a,b)$ such that*

$$f(x) = \int_I G(x,y)\,d\mu(y) + \frac{b-x}{b-a}f(a) + \frac{x-a}{b-a}f(b) \quad \text{for all } x \in [a,b] \tag{1.13}$$

and

$$\int_I (x-a)(b-x)\,d\mu(x) < \infty. \tag{1.14}$$

Proof. Consider first the case when f extends to a convex function in a neighborhood of $[a,b]$ (equivalently, when $f'_+(a)$ and $f'_-(b)$ exist and are finite). In this case we may choose as μ the Stieltjes measure associated to the nondecreasing function f'_+. In fact, integrating by parts we get

$$\begin{aligned}
\int_I G(x,y)\,d\mu(y) &= \int_I G(x,y)\,df'_+(y) \\
&= G(x,y)f'_+(y)\Big|_{y=a}^{y=b} - \int_I \frac{\partial G(x,y)}{\partial y} f'_+(y)\,dy \\
&= -\frac{x-b}{b-a}\int_a^x f'_+(y)\,dy - \frac{x-a}{b-a}\int_x^b f'_+(y)\,dy \\
&= -\frac{x-b}{b-a}(f(x)-f(a)) - \frac{x-a}{b-a}(f(b)-f(x)),
\end{aligned}$$

according to Proposition 1.6.1. This proves (1.13). Letting $x = (a+b)/2$ in (1.13) we get

$$\int_a^x (y-a)\,d\mu(y) + \int_x^b (b-y)\,d\mu(y) = f(a) + f(b) - 2f\Big(\frac{a+b}{2}\Big),$$

which yields

$$0 \leq \frac{1}{b-a}\int_I (x-a)(b-x)\,d\mu(x) \leq f(a) + f(b) - 2f\Big(\frac{a+b}{2}\Big). \tag{1.15}$$

In the general case we apply the above reasoning to the restriction of f to the interval $[a+\varepsilon, b-\varepsilon]$ and then pass to the limit as $\varepsilon \to 0$.

The uniqueness of μ is a consequence of the fact that $f'' = \mu$ in the sense of distribution theory. This can be easily checked by noticing that

$$\varphi(x) = \int_I G(x,y)\varphi''(y)\,dy$$

for all $\varphi \in C_c^2(I)$, which yields

$$\int_I f(x)\varphi''(x)\,dx = \iint_{I\times I} G(x,y)\varphi''(x)\,dx\,d\mu(y) = \int_I \varphi(y)\,d\mu(y),$$

due to the symmetry of G and the Fubini–Tonelli theorem. The application of this theorem was made possible by (1.14). $\square$

Theorem 1.6.3 shows that every continuous convex function on a compact interval is a superposition of an affine function and functions of the form $x \to G(x,y)$, equivalently, a superposition of an affine function and functions of the form $x \to (x-y)^+$ (or $x \to |x-y|$) for $y \in \mathbb{R}$. The essence of this fact was already noted at the end of Section 1.5.

Exercises

1. (The discrete analogue of Theorem 1.6.3) A sequence of real numbers $a_0, a_1, \ldots, a_n$ (with $n \geq 2$) is said to be *convex* provided that

$$\Delta^2 a_k = a_k - 2a_{k+1} + a_{k+2} \geq 0$$

for all $k = 0, \ldots, n-2$; it is said to be *concave* provided $\Delta^2 a_k \leq 0$ for all k.

(i) Solve the system

$$\Delta^2 a_k = b_k \quad \text{for } k = 0, \ldots, n-2$$

(in the unknowns a_k) to prove that the general form of a convex sequence $\mathbf{a} = (a_0, a_1, \ldots, a_n)$ with $a_0 = a_n = 0$ is given by the formula

$$\mathbf{a} = \sum_{j=1}^{n-1} c_j \mathbf{w}^j,$$

where $c_j = 2a_j - a_{j-1} - a_{j+1}$ and $\mathbf{w}^j$ has the components

$$w_k^j = \begin{cases} k(n-j)/n, & \text{for } k = 0, \ldots, j \\ j(n-k)/n, & \text{for } k = j, \ldots, n. \end{cases}$$

(ii) Prove that the general form of a convex sequence $\mathbf{a} = (a_0, a_1, \ldots, a_n)$ is $\mathbf{a} = \sum_{j=0}^{n} c_j \mathbf{w}^j$, where c_j and $\mathbf{w}^j$ are as in the case (i) for $j = 1, \ldots, n-1$. The other coefficients and components are:

$$c_0 = a_0,\ c_n = a_n,\ w_k^0 = (n-k)/n \text{ and } w_k^n = k/n \text{ for } k = 0, \ldots, n).$$

Remark. The theory of convex sequences can be subordinated to that of convex functions. If $f\colon [0,n] \to \mathbb{R}$ is a convex function, then $(f(k))_k$ is a convex sequence; conversely, if $(a_k)_k$ is a convex sequence, then the piecewise linear function $f\colon [0,n] \to \mathbb{R}$ obtained by joining the points (k, a_k) is convex too.

2. Prove the *discrete Berwald inequality*:

$$\frac{1}{n+1}\sum_{k=0}^{n} a_k \geq \Big(\frac{3(n-1)}{4(n+1)}\Big)^{1/2}\Big(\frac{1}{n+1}\sum_{k=0}^{n} a_k^2\Big)^{1/2}$$

for every concave sequence $a_0, a_1, \ldots, a_n$ of nonnegative numbers. [*Hint*: By Minkowski's inequality (Theorem 1.2.4 above), if the Berwald inequality works for two concave sequences, then it also works for all linear combinations of them, with positive coefficients. Then apply the assertion (ii) of the preceding exercise.]

1.7 Conjugate Convex Functions

The aim of this section is to develop a concept of duality between convex functions which makes possible an easy handling of some problems. The basic idea can be traced back to Young's inequality.

If $\varphi\colon [0,\infty) \to [0,\infty)$ is an increasing and continuous function with $\varphi(0) = 0$ and $\varphi(x) \to \infty$ as $x \to \infty$, then φ^{-1} exists and has the same properties as φ. Moreover, if we let

$$f(x) = \int_0^x \varphi(t)\,dt \quad \text{and} \quad f^*(y) = \int_0^y \varphi^{-1}(t)\,dt,$$

then f and f^* are both convex functions on $[0,\infty)$. By Young's inequality,

$$xy \leq f(x) + f^*(y) \quad \text{for all } x \geq 0,\ y \geq 0$$

(with equality if and only if $y = f'(x)$) and

$$f^*(y) = \sup\{xy - f(x) \mid x \geq 0\} \quad \text{for all } y \geq 0.$$

Clearly, the same is true if we extend f^* to $\mathbb{R}$ by letting $f^*(y) = 0$ for $y < 0$. Under these circumstances we say that f and f^* are *conjugate* functions.

If $\varphi(x) = x^{p-1}$, $x \in [0,\infty)$ (for $p > 1$), then $\varphi^{-1}(y) = y^{q-1}$, where $1/p + 1/q = 1$. In this case $f(x) = x^p/p$ and $f^*(y) = (y^+)^q/q$ for all $x \geq 0$, $y \in \mathbb{R}$.

The study of conjugate functions is important in several connections, for example, in the theory of Orlicz spaces. See Exercise 5.

In what follows we want to extend the notion of conjugacy to all convex functions, preserving its main features.

This is done by associating to each convex function $f\colon I \to \mathbb{R}$ defined on an interval I a new function,

$$f^*\colon I^* \to \mathbb{R}, \quad f^*(y) = \sup\{xy - f(x) \mid x \in I\},$$

with domain $I^* = \{y \in \mathbb{R} \mid f^*(y) < \infty\}$, called the *conjugate function* (or the *Legendre transform*) of f.

Lemma 1.7.1 *I^* is a nonempty interval and f^* is a convex function whose sublevel sets $L_\lambda = \{y \mid f^*(y) \leq \lambda\}$ are closed subsets of $\mathbb{R}$ for each $\lambda \in \mathbb{R}$.*

Proof. We first note that $I^* \neq \emptyset$. This is obvious if I is a singleton. If I is a nondegenerate interval, then for each $a \in \operatorname{int} I$ there is $y \in \mathbb{R}$ such that $f(x) \geq f(a) + y(x-a)$, which yields

$$xy - f(x) \leq ay - f(a),$$

so, $y \in I^*$. Our next remark is that I^* is actually an interval and f^* is a convex function. In fact, if $\lambda \in (0,1)$ and $y, z \in I^*$ then

$$\begin{aligned} f^*((1-\lambda)y + \lambda z) &= \sup\{x[(1-\lambda)y + \lambda z] - f(x) \mid x \in I\} \\ &\leq (1-\lambda)\sup\{xy - f(x) \mid x \in I\} \\ &\qquad + \lambda \sup\{xz - f(x) \mid x \in I\} \\ &= (1-\lambda)f^*(y) + \lambda f^*(z). \end{aligned}$$

It remains to prove that the sublevel sets $L_\lambda = \{y \mid f^*(y) \leq \lambda\}$ are closed. For this, consider a sequence $(y_n)_n$ of points of L_λ, which is converging, say to y. Then $xy_n - f(x) \leq f^*(y_n) \leq \lambda$ for each n and each x, so letting $n \to \infty$ we get $xy - f(x) \leq \lambda$ for each x, that is, $y \in I^*$ and $f^*(y) \leq \lambda$. $\square$

The real functions whose sublevel sets are closed are precisely the *lower semicontinuous* functions, that is, the functions $f\colon I \to \mathbb{R}$ such that

$$\liminf_{y \to x} f(y) = f(x)$$

at any point $x \in I$.

A moment's reflection shows that a convex function $f\colon I \to \mathbb{R}$ is lower semicontinuous if and only if it is continuous at each endpoint of I which belongs to I and $f(x) \to \infty$ as x approaches any finite endpoint not in I.

The following representation is a consequence of Proposition 1.6.1:

Lemma 1.7.2 *Let $f\colon I \to \mathbb{R}$ be a lower semicontinuous convex function and let φ be a real-valued function such that $\varphi(x) \in \partial f(x)$ for all $x \in I$. Then for all $a < b$ in I we have*

$$f(b) - f(a) = \int_a^b \varphi(t)\, dt.$$

We can now state the main result on the operation of conjugacy:

Theorem 1.7.3 *Let $f\colon I \to \mathbb{R}$ be a lower semicontinuous convex function. Then its conjugate $f^*\colon I^* \to \mathbb{R}$ is also convex and lower semicontinuous. Moreover:*

(i) *$xy \leq f(x) + f^*(y)$ for all $x \in I$, $y \in I^*$, with equality if and only if $y \in \partial f(x)$;*

(ii) $\partial f^* = (\partial f)^{-1}$ *as graphs of set-valued functions;*
(iii) $f^{**} = f$.

Recall that the inverse of a graph G is the set $G^{-1} = \{(y,x) \mid (x,y) \in G\}$.

Proof. The first assertion follows from Lemma 1.7.1.

(i) The inequality is immediate. To settle the equality case, we need the fact that a convex function h attains a minimum at x if and only if $\partial f(x)$ contains 0 (see Section 1.5, Exercise 1 (ii)). Since the function $h(x) = f(x) - xy$ is convex and

$$-f^*(y) = -\sup\{xy - f(x) \mid x \in I\} = \inf\{f(x) - xy \mid x \in I\},$$

the equality $-f^*(y) = f(x) - xy$ holds true if and only if $0 \in \partial h(x)$, that is, if $y \in \partial f(x)$.

(ii) Letting $x \in I$, we have $f^*(y) - xy \geq -f(x)$ for all $y \in I^*$. Thus $f^*(y) - xy = -f(x)$ occurs at a point of minimum, which by the above discussion means $y \in \partial f(x)$. In other words, $y \in \partial f(x)$ implies that the function $h(z) = f^*(z) - xz$ has a (global) minimum at $z = y$. Since h is convex, this could happen only if $0 \in \partial h(y)$, that is, when $x \in \partial(f^*)(y)$. Taking inverses, we are led to

$$x \in (\partial f)^{-1}(y) \implies x \in \partial(f^*)(y),$$

which shows that $\partial(f^*) \supset (\partial f)^{-1}$. For equality we remark that the graph G of the subdifferential of any lower semicontinuous convex function f is maximal monotone in the sense that

$$(x_1, y_1) \text{ and } (x_2, y_2) \text{ in } G \implies (x_2 - x_1)(y_2 - y_1) \geq 0 \tag{1.16}$$

and G is not a proper subset of any other subset of $\mathbb{R} \times \mathbb{R}$ with the property (1.16). Then the graph of $(\partial f)^{-1}$ is maximal monotone (being the inverse of a maximal monotone set) and thus $(\partial f)^{-1} = \partial(f^*)$, which is (ii).

The implication (1.16) follows easily from the monotonicity of the subdifferential, while maximality can be established by *reductio ad absurdum.*

(iii) According to (ii), $\partial(f^{**}) = (\partial(f^*))^{-1} = (\partial(f)^{-1})^{-1} = \partial f$, which yields that $I = I^{**}$, and so by Lemma 1.7.2,

$$f(x) - f(c) = \int_c^x \varphi(t)\, dt = f^{**}(x) - f^{**}(c)$$

for all $x, c \in I$. It remains to find a $c \in I$ for which $f(c) = f^{**}(c)$. Choose $z \in I$ and $y \in I^*$ such that $y \in \partial f(z)$. By (iii), this means $z \in (\partial f)^{-1}(y) = \partial f^*(y)$. According to (i), applied for f and f^*, we have

$$zy = f(z) + f^*(y) = f^*(y) + f^{**}(z),$$

that is, $f(z) = f^{**}(z)$. $\square$

By Theorem 1.7.3, if f is differentiable, then its conjugate can be determined by eliminating x from the equations

$$f(x) + f^*(y) = xy \quad \text{and} \quad f'(x) = y.$$

Exercises

1. Let f be a lower semicontinuous convex function defined on a bounded interval I. Prove that $I^* = \mathbb{R}$.

2. Compute ∂f, ∂f^* and f^* for $f(x) = |x|$, $x \in \mathbb{R}$.

3. Prove that:

(i) the conjugate of $f(x) = |x|^p/p$, $x \in \mathbb{R}$, is

$$f^*(y) = |y|^q/q, \quad y \in \mathbb{R} \ (p > 1, \frac{1}{p} + \frac{1}{q} = 1);$$

(ii) the conjugate of $f(x) = (1 + x^2)^{1/2}$, $x \in \mathbb{R}$, is the function

$$f^*(y) = -(1 - y^2)^{1/2}, \quad y \in [-1, 1];$$

(iii) the conjugate of $f(x) = \mathrm{e}^x$, $x \in \mathbb{R}$, is the function $f^*(y) = y \log y - y$ for $y > 0$ and $f^*(0) = 0$;

(iv) the conjugate of $f(x) = -\log x$, $x > 0$, is the function

$$f^*(y) = -1 - \log(-y), \quad y < 0.$$

4. (A minimization problem) Let $f \colon \mathbb{R} \to \mathbb{R}$ be a convex function such that $\lim_{|x| \to \infty} f(x) = \infty$. Consider the function

$$F(x) = \inf \Big\{ \sum_{k=1}^{n} c_k f(x_k) \mid x_1, \ldots, x_n \in \mathbb{R}, \ \sum_{k=1}^{n} c_k a_k x_k = x \Big\}$$

where $c_1, \ldots, c_n$ are given positive constants and $a_1, \ldots, a_n$ are given nonzero constants. Prove that F is convex and that its conjugate is the function

$$F^*(y) = \sum_{k=1}^{n} c_k f^*(a_k y).$$

5. An *Orlicz function* is any convex function $\Phi \colon [0, \infty) \to \mathbb{R}$ such that:

(Φ1) $\Phi(0) = 0$, $\Phi(x) > 0$ for $x > 0$;
(Φ2) $\Phi(x)/x \to 0$ as $x \to 0$ and $\Phi(x)/x \to \infty$ as $x \to \infty$;
(Φ3) there exists a positive constant K such that $\Phi(2x) \leq K\Phi(x)$ for $x \geq 0$.

Let (X, Σ, μ) be a complete σ-finite measure space and let $S(\mu)$ be the vector space of all equivalence classes of μ-measurable real-valued functions defined on X. The *Orlicz space* $L^\Phi(X)$ is the subspace of all $f \in S(\mu)$ such that

$$I_\Phi(f/\lambda) = \int_X \Phi(|f(x)|/\lambda)\, d\mu < \infty, \quad \text{for some } \lambda > 0.$$

(i) Prove that $L^\Phi(X)$ is a linear space such that

$$|f| \leq |g| \quad \text{and} \quad g \in L^\Phi(X) \text{ imply } f \in L^\Phi(X).$$

(ii) Prove that $L^\Phi(X)$ is a Banach space when endowed with the norm

$$\|f\|_\Phi = \inf\{\lambda > 0 \mid I_\Phi(f/\lambda) \leq 1\}.$$

(iii) Prove that the dual of $L^\Phi(X)$ is $L^\Psi(X)$, where Ψ is the conjugate of the function Φ.

Remark. The Orlicz spaces extend the $L^p(\mu)$ spaces. Their theory is exposed in books like [132] and [209]. The Orlicz space $\mathcal{L}\log^+\mathcal{L}$ (corresponding to the Lebesgue measure on $[0,\infty)$ and the function $\Phi(t) = t(\log t)^+$) plays a role in Fourier analysis. See [253]. Applications to interpolation theory are given in [20].

1.8 The Integral Form of Jensen's Inequality

The analogue of the arithmetic mean in the context of finite measure spaces (X, Σ, μ) is the *integral arithmetic mean* (or, simply, the *arithmetic mean*), which, for a μ-integrable function $f\colon X \to \mathbb{R}$, is the number

$$M_1(f;\mu) = \frac{1}{\mu(X)} \int_X f\, d\mu.$$

For convenience, we shall denote $M_1(f;\mu)$ also $M_1(f)$.

In probability theory, $M_1(f)$ represents the *conditional expectation* of the random variable f (in which case it is denoted $\mathcal{E}(f)$).

There are many results on the integral arithmetic mean. A basic one is the integral form of Jensen's inequality:

Theorem 1.8.1 (Jensen's inequality) *Let (X, Σ, μ) be a finite measure space and let $g\colon X \to \mathbb{R}$ be a μ-integrable function. If f is a convex function given on an interval I that includes the image of g, then $M_1(g) \in I$ and*

$$f(M_1(g)) \leq M_1(f \circ g)$$

provided that $f \circ g$ is μ-integrable.

If f is strictly convex, then the above inequality becomes an equality if and only if g is constant μ-almost everywhere.

Proof. $M_1(g)$ belongs to I since otherwise $h = M_1(g) - g$ (or $-h$) will be a strictly positive function whose integral is 0.

Then, choose a function $\varphi\colon I \to \mathbb{R}$ such that $\varphi(x) \in \partial f(x)$ for all $x \in \operatorname{int} I$. If $M_1(g) \in \operatorname{int} I$, then

$$f(g(x)) \geq f(M_1(g)) + (g(x) - M_1(g)) \cdot \varphi(M_1(g)) \quad \text{for all } x \in X$$

and Jensen's inequality follows by integrating both sides over X. The case where $M_1(g)$ is an endpoint of I is straightforward because in that case $g = M_1(g)$ μ-almost everywhere. $\square$

Remark 1.8.2 (The integral form of the arithmetic-geometric-harmonic mean inequality) Consider a finite measure space (X, Σ, μ) and a function $f \in L^1(\mu)$ such that $f \geq 0$. Define $\log 0 = -\infty$ and $\mathrm{e}^{-\infty} = 0$. According to Jensen's inequality,

$$\frac{1}{\mu(X)} \int_X \log f(x)\, d\mu \leq \log\left(\frac{1}{\mu(X)} \int_X f(x)\, d\mu\right),$$

the inequality being strict except for the case when f is a constant function μ-almost everywhere. This fact can be restated as $M_0(f;\mu) \leq M_1(f;\mu)$, where

$$M_0(f;\mu) = \exp\left(\frac{1}{\mu(X)} \int_X \log f(x)\, d\mu\right)$$

represents the *geometric mean* of f. If we agree to regard 0 and ∞ as reciprocals of one another, we may introduce also the *harmonic mean* of f,

$$M_{-1}(f;\mu) = \left(\frac{1}{\mu(X)} \int_X \frac{1}{f(x)}\, d\mu\right)^{-1}.$$

It is clear that $M_0(f;\mu) = (M_0(1/f;\mu))^{-1}$ and $M_{-1}(f;\mu) = (M_1(1/f;\mu))^{-1}$, so that

$$M_{-1}(f;\mu) \leq M_0(f;\mu) \leq M_1(f;\mu).$$

In Section 3.6 we shall prove that Jensen's inequality still works (under additional hypotheses) outside the framework of finite measure spaces.

Jensen's inequality can be related to another well-known inequality:

Theorem 1.8.3 (Chebyshev's inequality) *If $g, h\colon [a,b] \to \mathbb{R}$ are Riemann integrable and synchronous* (*in the sense that*

$$(g(x) - g(y))(h(x) - h(y)) \geq 0$$

for all $x, y \in [a,b]$), *then $M_1(g)M_1(h) \leq M_1(gh)$.*

The next result complements Jensen's inequality by Chebyshev's inequality:

Theorem 1.8.4 (The complete form of Jensen's inequality) *Let (X, Σ, μ) be a finite measure space and let $g\colon X \to \mathbb{R}$ be a μ-integrable function. If f is a convex function given on an interval I that includes the image of g and $\varphi\colon I \to \mathbb{R}$ is a function such that*

(i) $\varphi(x) \in \partial f(x)$ *for every $x \in I$, and*
(ii) $\varphi \circ g$ *and $g \cdot (\varphi \circ g)$ are μ-integrable functions,*

then the following inequalities hold:

$$0 \leq M_1(f \circ g) - f(M_1(g)) \leq M_1(g \cdot (\varphi \circ g)) - M_1(g) M_1(\varphi \circ g).$$

If f is concave, then the inequalities in Theorems 1.8.1 and 1.8.4 hold in the reversed direction.

Proof. The first inequality is that of Jensen. The second can be obtained from

$$f(M_1(g)) \geq f(g(x)) + (M_1(g) - g(x)) \cdot \varphi(g(x)) \quad \text{for all } x \in X$$

by integrating both sides over X. □

The following result represents a discrete version of Theorem 1.8.4:

Corollary 1.8.5 *Let f be a convex function defined on an open interval I and let $\varphi\colon I \to \mathbb{R}$ be a function such that $\varphi(x) \in \partial f(x)$ for all $x \in I$. Then*

$$\begin{aligned} 0 &\leq \sum_{k=1}^{n} \lambda_k f(x_k) - f\Big(\sum_{k=1}^{n} \lambda_k x_k\Big) \\ &\leq \sum_{k=1}^{n} \lambda_k x_k \varphi(x_k) - \Big(\sum_{k=1}^{n} \lambda_k x_k\Big)\Big(\sum_{k=1}^{n} \lambda_k \varphi(x_k)\Big) \end{aligned}$$

for all $x_1, \ldots, x_n \in I$ and all $\lambda_1, \ldots, \lambda_n \in [0,1]$, with $\sum_{k=1}^{n} \lambda_k = 1$.

An application of Corollary 1.8.5 is indicated in Exercise 6.

In a symmetric way, we may complement Chebyshev's inequality by Jensen's inequality:

Theorem 1.8.6 (The complete form of Chebyshev's inequality) *Let (X, Σ, μ) be a finite measure space, let $g\colon X \to \mathbb{R}$ be a μ-integrable function and let φ be a nondecreasing function given on an interval that includes the image of g and such that $\varphi \circ g$ and $g \cdot (\varphi \circ g)$ are integrable functions. Then for every primitive Φ of φ for which $\Phi \circ g$ is integrable, the following inequalities hold true:*

$$0 \leq M_1(\Phi \circ g) - \Phi(M_1(g)) \leq M_1(g \cdot (\varphi \circ g)) - M_1(g) M_1(\varphi \circ g).$$

In order to show how Theorem 1.8.6 yields Chebyshev's inequality we will consider first the case when $g\colon [a,b] \to \mathbb{R}$ is increasing and $h\colon [a,b] \to \mathbb{R}$ is nondecreasing. In this case we apply Theorem 1.8.6 to g and $\varphi = h \circ g^{-1}$. When both g and h are nondecreasing, we consider increasing perturbations of g, for example, $g + \varepsilon x$ with $\varepsilon > 0$. By the previous case,

$$M_1(g + \varepsilon x)M_1(h) \le M_1((g + \varepsilon x)h)$$

and Chebyshev's inequality follows by taking the limit as $\varepsilon \to 0+$.

In connection with Jensen's inequality, it is important to notice here another classical inequality:

Theorem 1.8.7 (Hardy's inequality) *Suppose that $f \in L^p(0,\infty)$, $f \ge 0$, where $p \in (1,\infty)$. Put*

$$F(x) = \frac{1}{x}\int_0^x f(t)\,dt, \quad x > 0.$$

Then

$$\|F\|_{L^p} \le \frac{p}{p-1}\,\|f\|_{L^p}$$

with equality if and only if $f = 0$ almost everywhere.

Hardy's inequality yields the norm of the averaging operator $H\colon f \to F$, from $L^p(0,\infty)$ into $L^p(0,\infty)$. In fact, the constant $p/(p-1)$ is best possible (though untainted). The optimality can easily be checked by considering the sequence of functions $f_\varepsilon(t) = (t^{-1/p} + \varepsilon)\chi_{(0,1]}(t)$, and letting $\varepsilon \to 0+$.

Hardy's inequality can be deduced from the following lemma:

Lemma 1.8.8 *Let $0 < b < \infty$ and $-\infty \le a < c \le \infty$. If u is a positive convex function on (a,c), then*

$$\int_0^b u\Big(\frac{1}{x}\int_0^x h(t)\,dt\Big)\frac{dx}{x} \le \int_0^b u(h(x))\Big(1 - \frac{x}{b}\Big)\frac{dx}{x}$$

for all integrable functions $h\colon (0,b) \to (a,c)$.

Proof. In fact, by Jensen's inequality,

$$\begin{aligned}
\int_0^b u\Big(\frac{1}{x}\int_0^x h(t)\,dt\Big)\frac{dx}{x} &\le \int_0^b \Big(\frac{1}{x}\int_0^x u(h(t))\,dt\Big)\frac{dx}{x}\\
&= \int_0^b \frac{1}{x^2}\Big(\int_0^b u(h(t))\chi_{[0,x]}(t)\,dt\Big)\,dx\\
&= \int_0^b u(h(t))\Big(\int_t^b \frac{1}{x^2}\,dx\Big)\,dt\\
&= \int_0^b u(h(t))\Big(1 - \frac{t}{b}\Big)\frac{dt}{t}
\end{aligned}$$

and the proof is complete. □

For $u(x) = |x|^p$, the result of Lemma 1.8.8 can be put in the following form

$$\int_0^\alpha \left|\frac{1}{x}\int_0^x f(t)\,dt\right|^p dx \le \left(\frac{p}{p-1}\right)^p \int_0^\alpha |f(x)|^p \left(1 - \left(\frac{x}{\alpha}\right)^{(p-1)/p}\right) dx, \quad (1.17)$$

where $\alpha = b^{p/(p-1)}$ and $f(x) = h(x^{1-1/p})x^{-1/p}$. This yields an analogue of Hardy's inequality for functions $f \in L^p(0,\alpha)$ (where $0 < \alpha < \infty$), from which Hardy's inequality follows by letting $\alpha \to \infty$.

The equality case in Theorem 1.8.7 implies that F and f are proportional, which makes f of the form Cx^r. Since $f \in L^p(0,\infty)$, this is possible only for $C = 0$.

As shown in the above argument, Hardy's inequality also holds for $p < 0$.

An alternative proof of Theorem 1.8.7 can be built on the well-known fact that $C_c(0,\infty)$ (the space of all continuous functions $f\colon (0,\infty) \to \mathbb{C}$ with compact support) is dense in $L^p(0,\infty)$. This allows us to restrict ourselves to the case where $f \in C_c(0,\infty)$. Then $(xF(x))' = f(x)$, which yields

$$\begin{aligned}\int_0^\infty F^p(t)\,dt &= -p\int_0^\infty F^{p-1}(t)tF'(t)\,dt \\ &= -p\int_0^\infty F^{p-1}(t)(f(t) - F(t))\,dt,\end{aligned}$$

that is, $\int_0^\infty F^p(t)\,dt = \frac{p}{p-1}\int_0^\infty F^{p-1}(t)f(t)\,dt$. The proof ends by taking into account the integral form of the Rogers–Hölder inequality.

Exercises

1. (The power means; see Section 1.1, Exercise 8, for the discrete case) Consider a finite measure space (X,Σ,μ). The *power mean* of order $t \ne 0$ is defined for all nonnegative measurable functions $f\colon X \to \mathbb{R}$, $f^t \in L^1(\mu)$, by the formula

$$M_t(f\,;\mu) = \left(\frac{1}{\mu(X)}\int_X f^t\,d\mu\right)^{1/t}.$$

We also define

$$M_0(f\,;\mu) = \exp\left(\frac{1}{\mu(X)}\int_X \log f(x)\,d\mu\right)$$

for $f \in \bigcup_{t>0} L^t(\mu)$, $f \ge 0$, and

$$\begin{aligned}M_{-\infty}(f\,;\mu) &= \sup\bigl\{\alpha \ge 0 \bigm| \mu(\{x \in X \mid f(x) < \alpha\}) = 0\bigr\} \\ M_\infty(f\,;\mu) &= \inf\bigl\{\alpha \ge 0 \bigm| \mu(\{x \in X \mid f(x) > \alpha\}) = 0\bigr\}\end{aligned}$$

for $f \in L^\infty(\mu)$, $f \ge 0$.

(i) (Jensen's inequality for means) Suppose that $-\infty \leq s \leq t \leq \infty$ and $M_t(f\,;\mu) < \infty$. Prove that

$$M_s(f\,;\mu) \leq M_t(f;\mu).$$

(ii) Suppose that $f \in L^\infty(\mu)$, $f \geq 0$. Prove that

$$\lim_{t\to-\infty} M_t(f\,;\mu) = M_{-\infty}(f;\mu) \quad \text{and} \quad \lim_{t\to\infty} M_t(f\,;\mu) = M_\infty(f;\mu).$$

(iii) Suppose that $f \in L^\infty(\mu)$, $f \geq 0$. Prove the convexity of the function $t \mapsto t \log M_t(f\,;\mu)$ on $\mathbb{R}$.

(iv) Notice that $(t^r - 1)/r$ decreases to $\log t$ as $r \downarrow 0$ and apply the dominated convergence theorem of Lebesgue to conclude that

$$\lim_{r\to 0+} M_r(f;\mu) = M_0(f\,;\mu)$$

for all $f \in L^1(\mu)$, $f \geq 0$.

2. Infer from the integral form of the arithmetic-geometric-harmonic mean inequality that $L(a,b) < I(a,b) < A(a,b)$ for all $a, b > 0$, $a \neq b$. Here $L(a,b)$, $I(a,b)$, $A(a,b)$ are the logarithmic, the identric and respectively the arithmetic mean of a and b.

3. Infer from Theorem 1.8.7 the discrete form of Hardy's inequality:

$$\Big(\sum_{n=1}^\infty \Big(\frac{1}{n}\sum_{k=1}^n a_k\Big)^p\Big)^{1/p} < \frac{p}{p-1}\Big(\sum_{k=1}^\infty a_k^p\Big)^{1/p},$$

for every sequence $(a_n)_n$ of nonnegative numbers (not all zero) and every $p \in (1,\infty)$.

4. (The Pólya–Knopp inequality; see [99], [130]) Prove the following limiting case of Hardy's inequality: for every $f \in L^1(0,\infty)$, $f \geq 0$ and f not identically zero,

$$\int_0^\infty \exp\Big(\frac{1}{x}\int_0^x \log f(t)\,dt\Big)\,dx < e\int_0^\infty f(x)\,dx.$$

The discrete form of this inequality was previously noted by T. Carleman [48]:

$$\sum_{n=1}^\infty (a_1 a_2 \cdots a_n)^{1/n} < e\sum_{n=1}^\infty a_n$$

for $a_1, a_2, a_3, \ldots \geq 0$, not all zero.
[*Hint*: Apply Lemma 1.8.8 for $h(x) = \log f(x)$.]

5. Formulate and prove the analogue of the Pólya–Knopp inequality for functions defined on bounded intervals of the form $(0,\alpha)$.

6. Assume that A, B, C are the angles of a triangle (expressed in radians). Prove that

$$0 \leq 3\sqrt{3}/2 - \sum \sin A \leq \sum \Big(\frac{\pi}{3} - A\Big) \cos A.$$

7. (Another estimate of Jensen's inequality) Let (X, Σ, μ) be a finite measure space and let $g \in L^2(\mu)$. If f is a twice differentiable function given on an interval I that includes the image of g and $\alpha \leq f''/2 \leq \beta$, then

$$\alpha \operatorname{var}(g) \leq M_1(f \circ g) - f(M_1(g)) \leq \beta \operatorname{var}(g).$$

Here $\operatorname{var}(g) = M_1\big((g - M_1(g))^2\big)$ denotes the *variance* of g.
[*Hint*: Apply Taylor's formula to infer that

$$\begin{aligned}\alpha(g(x) - M_1(g))^2 &\leq f(g(x)) - f(M_1(g)) - f'(M_1(g))(g(x) - M_1(g)) \\ &\leq \beta(g(x) - M_1(g))^2. \quad]\end{aligned}$$

1.9 The Hermite–Hadamard Inequality

As stated in Proposition 1.3.4 above, every convex function f on an interval $[a, b]$ can be modified at the endpoints to become convex and continuous. An immediate consequence of this fact is the (Riemann) integrability of f. The arithmetic mean of f can be estimated by the *Hermite–Hadamard inequality*,

$$f\Big(\frac{a+b}{2}\Big) \leq \frac{1}{b-a}\int_a^b f(x)\,dx \leq \frac{f(a)+f(b)}{2}. \tag{1.18}$$

The right-hand side (denoted (RHH)) follows by integrating the inequality (1.2) in Section 1.1 (which says that the graph is under the chord joining the endpoints). Assuming that f is also continuous, we actually get

$$\frac{1}{b-a}\int_a^b f(x)\,dx < \frac{f(a)+f(b)}{2}$$

except when f is *affine*, that is, when $f(x) = f(a) + \frac{f(b)-f(a)}{b-a}(x-a)$.

The left-hand side of (1.18) (denoted (LHH)) is also easy to prove:

$$\begin{aligned}\frac{1}{b-a}\int_a^b f(x)\,dx &= \frac{1}{b-a}\Big(\int_a^{(a+b)/2} f(x)\,dx + \int_{(a+b)/2}^b f(x)\,dx\Big) \\ &= \frac{1}{2}\int_0^1 \Big[f\Big(\frac{a+b-t(b-a)}{2}\Big) + f\Big(\frac{a+b+t(b-a)}{2}\Big)\Big]\,dt \\ &\geq f\Big(\frac{a+b}{2}\Big).\end{aligned}$$

Interestingly, each of the two sides of (1.18) in fact characterizes convex functions. More precisely, *if I is an interval and $f\colon I \to \mathbb{R}$ is a continuous* *function whose restriction to every compact subinterval* $[a, b]$ *verifies* (LHH)*, then* f *is convex.* The same works when (LHH) is replaced by (RHH). See Exercises 1 and 2.

We shall illustrate the power of the Hermite–Hadamard inequality with several examples from calculus.

Examples 1.9.1

(i) For $f(x) = 1/(1+x)$, $x \geq 0$, Ch. Hermite [102] observed that

$$x - x^2/(2+x) < \log(1+x) < x - x^2/(2+2x).$$

Particularly,

$$\frac{1}{n+1/2} < \log(n+1) - \log n < \frac{1}{2}\Big(\frac{1}{n} + \frac{1}{n+1}\Big) \tag{1.19}$$

for all $n \in \mathbb{N}^*$, and this fact is instrumental in deriving Stirling's formula,

$$n! \sim \sqrt{2\pi} \cdot n^{n+1/2} \mathrm{e}^{-n}.$$

See Theorem 2.2.11.

(ii) For $f = \exp$, the inequality (1.18) yields

$$\mathrm{e}^{(a+b)/2} < \frac{\mathrm{e}^b - \mathrm{e}^a}{b-a} < \frac{\mathrm{e}^a + \mathrm{e}^b}{2} \quad \text{for } a \neq b \text{ in } \mathbb{R},$$

that is,

$$\sqrt{xy} < \frac{x-y}{\log x - \log y} < \frac{x+y}{2} \quad \text{for } x \neq y \text{ in } (0, \infty), \tag{1.20}$$

which represents the geometric–logarithmic–arithmetic mean inequality. For $f = \log$, we obtain a similar inequality, where the role of the logarithmic mean is taken by the identric mean.

(iii) For $f(x) = \sin x$, $x \in [0, \pi]$, we obtain

$$\frac{\sin a + \sin b}{2} < \frac{\cos a - \cos b}{b-a} < \sin\Big(\frac{a+b}{2}\Big) \quad \text{for } a \neq b \text{ in } \mathbb{R},$$

and this implies the well-known inequalities $\tan x > x > \sin x$ (for x in $(0, \pi/2)$).

The following result yields an estimate of the precision in the Hermite–Hadamard inequality:

Lemma 1.9.2 *Let* $f\colon [a, b] \to \mathbb{R}$ *be a twice differentiable function for which there exist real constants* m *and* M *such that*

$$m \leq f'' \leq M.$$

Then

$$m \cdot \frac{(b-a)^2}{24} \le \frac{1}{b-a}\int_a^b f(x)\,dx - f\Big(\frac{a+b}{2}\Big) \le M \cdot \frac{(b-a)^2}{24},$$

and

$$m \cdot \frac{(b-a)^2}{12} \le \frac{f(a)+f(b)}{2} - \frac{1}{b-a}\int_a^b f(x)\,dx \le M \cdot \frac{(b-a)^2}{12}.$$

Proof. In fact, the functions $f - mx^2/2$ and $Mx^2/2 - f$ are convex and thus we can apply to them the Hermite–Hadamard inequality. □

For other estimates see the Comments at the end of this chapter.

Remark 1.9.3 (An improvement on the Hermite–Hadamard inequality) Suppose that $f\colon [a,b] \to \mathbb{R}$ is a convex function. By applying the Hermite–Hadamard inequality on each of the intervals $[a,(a+b)/2]$ and $[(a+b)/2,b]$ we get

$$f\Big(\frac{3a+b}{4}\Big) \le \frac{2}{b-a}\int_a^{(a+b)/2} f(x)\,dx \le \frac{1}{2}\Big(f(a) + f\Big(\frac{a+b}{2}\Big)\Big)$$

and

$$f\Big(\frac{a+3b}{4}\Big) \le \frac{2}{b-a}\int_{(a+b)/2}^b f(x)\,dx \le \frac{1}{2}\Big(f\Big(\frac{a+b}{2}\Big) + f(b)\Big).$$

Summing up (side by side), we obtain the following refinement of (1.18):

$$\begin{aligned} f\Big(\frac{a+b}{2}\Big) &\le \frac{1}{2}\Big(f\Big(\frac{3a+b}{4}\Big) + f\Big(\frac{a+3b}{4}\Big)\Big) \\ &\le \frac{1}{b-a}\int_a^b f(x)\,dx \le \frac{1}{2}\Big[f\Big(\frac{a+b}{2}\Big) + \frac{f(a)+f(b)}{2}\Big] \\ &\le \frac{1}{2}(f(a)+f(b)). \end{aligned}$$

By continuing the division process, the arithmetic mean of f can be approximated as close as desired by convex combinations of values of f at suitable dyadic points of $[a,b]$.

The Hermite–Hadamard inequality is the starting point to Choquet's theory, which is the subject of Chapter 4.

Exercises

1. Infer from Theorem 1.1.3 that a (necessary and) sufficient condition for a continuous function f to be convex on an open interval I is that

$$f(x) \leq \frac{1}{2h} \int_{x-h}^{x+h} f(t)\, dt$$

for all x and h with $[x-h, x+h] \subset I$.

2. Let f be a real-valued continuous function defined on an open interval I. Prove that f is convex if it verifies the right-hand side inequality in (1.18).

3. (An improvement of the left-hand side of (1.18)) Let $f\colon [a,b] \to \mathbb{R}$ be a convex function. Use the existence of support lines to show that

$$\frac{1}{2}\Big(f\Big(\frac{a+b}{2} - c\Big) + f\Big(\frac{a+b}{2} + c\Big)\Big) \leq \frac{1}{b-a} \int_a^b f(x)\, dx$$

for all $c \in [0, (b-a)/4]$, and that $c = (b-a)/4$ is maximal within the class of convex functions on $[a,b]$.

4. Notice the following quadrature formula,

$$\frac{1}{b-a} \int_a^b f(x)\, dx = \frac{1}{2}\,[f(a) + f(b)] - \frac{1}{b-a} \int_a^b f''(x) \frac{(b-x)(x-a)}{2}\, dx,$$

valid for $f \in C^2([a,b], \mathbb{R})$, and infer from it the right-hand side of inequality (1.18).

5. (Pólya's inequality) Prove that

$$\frac{x-y}{\log x - \log y} < \frac{1}{3}\Big(2\sqrt{xy} + \frac{x+y}{2}\Big) \quad \text{for all } x \neq y \text{ in } (0, \infty).$$

1.10 Convexity and Majorization

In a celebrated paper published in 1929, G. H. Hardy, J. E. Littlewood and G. Pólya [98] proved an interesting characterization of convex functions in terms of an order relation defined on $\mathbb{R}^n$, called by them *majorization*. Their basic observation was the subject of Theorem 1.5.4 above.

For any vector $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, let

$$x_1^{\downarrow} \geq \cdots \geq x_n^{\downarrow}$$

denote the components of x in decreasing order. For $x, y \in \mathbb{R}^n$ we put $x \prec y$ (and say that x is *majorized* by y, or that y *majorizes* x) if

$$\sum_{i=1}^{k} x_i^{\downarrow} \leq \sum_{i=1}^{k} y_i^{\downarrow} \quad \text{for } k = 1, \dots, n-1$$

$$\sum_{i=1}^{n} x_i^{\downarrow} = \sum_{i=1}^{n} y_i^{\downarrow}.$$

The geometric insight into majorization was later observed by R. Rado [208]: $x \prec y$ means that the components of x spread out less than those of y in the sense that x lies in the convex hull of the $n!$ permutations of y (the orbit of y under the group of permutation matrices). See Exercise 6, Section 3.3.

This fact comes from another characterization of the majorization relation via doubly stochastic matrices. Recall that a matrix $P \in \mathrm{M}_n(\mathbb{R})$ is *doubly stochastic* if P has nonnegative entries and each row and each column sums to unity. A special class of doubly stochastic matrices is that of *T-transformations*. They have the form

$$T = \lambda I + (1 - \lambda)Q$$

where $0 \leq \lambda \leq 1$ and Q is a permutation mapping which interchanges two coordinates, that is,

$$Tx = \big(x_1, \ldots, x_{j-1}, \lambda x_j + (1-\lambda)x_k, x_{j+1}, \ldots, x_{k-1}, \lambda x_k \\ + (1-\lambda)x_j, x_{k+1}, \ldots, x_n\big).$$

Theorem 1.10.1 *Let $x, y \in \mathbb{R}^n$. Then the following assertions are equivalent:*

(i) $x \prec y$;

(ii) $\sum_k f(x_k) \leq \sum_k f(y_k)$ *for every continuous convex function f whose domain of definition contains the components of x and y*;

(iii) $\sum_\pi \alpha_{\pi(1)}^{x_1} \cdots \alpha_{\pi(n)}^{x_n} \leq \sum_\pi \alpha_{\pi(1)}^{y_1} \cdots \alpha_{\pi(n)}^{y_n}$ *for every $\alpha_1, \ldots, \alpha_n > 0$, the sum being taken over all permutations π of the set* $\{1, \ldots, n\}$;

(iv) $x = Py$ *for a suitable doubly stochastic matrix* $P \in \mathrm{M}_n(\mathbb{R})$;

(v) *x can be obtained from y by successive applications of finitely many T-transformations.*

The equivalence (i) $\Leftrightarrow$ (ii) is due to G. H. Hardy, J. E. Littlewood and G. Pólya [98]. The implication (iv) $\Rightarrow$ (i) is due to I. Schur [224]. The equivalence (i) $\Leftrightarrow$ (iii) is due to R. F. Muirhead [172], for x and y having nonnegative integer components. The implication (i) $\Rightarrow$ (v) was first observed by Muirhead in the case of points with integer components and later extended by Hardy, Littlewood and Pólya to the general case.

Proof. (i) $\Rightarrow$ (ii) See Theorem 1.5.4 above.

(ii) $\Rightarrow$ (i) Since the identity and its opposite are convex functions, we get $\sum_{i=1}^n x_i^{\downarrow} = \sum_{i=1}^n y_i^{\downarrow}$. Also, using the convexity of $f = (x - y_k^{\downarrow})^+$, we get

$$x_1^{\downarrow} + \cdots + x_k^{\downarrow} - k y_k^{\downarrow} \leq \sum_{j=1}^k f(x_j^{\downarrow}) \leq \sum_{j=1}^k f(y_j^{\downarrow}) \leq y_1^{\downarrow} + \cdots + y_k^{\downarrow} - k y_k^{\downarrow}$$

that is, $x_1^{\downarrow} + \cdots + x_k^{\downarrow} \leq y_1^{\downarrow} + \cdots + y_k^{\downarrow}$.

(i) $\Rightarrow$ (iii) It suffices to consider the case where x and y differ in only two components, say $x_k = y_k$ for $k \geq 3$. Relabel if necessary, so that $x_1 > x_2$ and $y_1 > y_2$. Then there exists $\delta > 0$ such that $y_1 = x_1 + \delta$ and $y_2 = x_2 - \delta$. We have

$$\begin{aligned}
&\sum_{\pi} \alpha_{\pi(1)}^{y_1} \cdots \alpha_{\pi(n)}^{y_n} - \sum_{\pi} \alpha_{\pi(1)}^{x_1} \cdots \alpha_{\pi(n)}^{x_n} \\
&\quad = \frac{1}{2} \sum_{\pi} \Big[\alpha_{\pi(1)}^{y_1} \alpha_{\pi(2)}^{y_2} - \alpha_{\pi(1)}^{y_1-\delta} \alpha_{\pi(2)}^{y_2+\delta} + \alpha_{\pi(2)}^{y_1} \alpha_{\pi(1)}^{y_2} - \alpha_{\pi(2)}^{y_1-\delta} \alpha_{\pi(1)}^{y_2+\delta} \Big] \prod_{k=3}^{n} \alpha_{\pi(k)}^{y_k} \\
&\quad = \frac{1}{2} \sum_{\pi} (\alpha_{\pi(1)} \alpha_{\pi(2)})^{y_2} \big(\alpha_{\pi(1)}^{y_1-y_2-\delta} - \alpha_{\pi(2)}^{y_1-y_2-\delta} \big) (\alpha_{\pi(1)}^{\delta} - \alpha_{\pi(2)}^{\delta}) \prod_{k=3}^{n} \alpha_{\pi(k)}^{y_k} \\
&\quad \geq 0.
\end{aligned}$$

(iii) $\Rightarrow$ (i) The case where $\alpha_1 = \cdots = \alpha_n > 0$ gives us

$$\alpha_1^{\sum_{k=1}^n x_k} \leq \alpha_1^{\sum_{k=1}^n y_k}$$

so that $\sum_{k=1}^n x_k = \sum_{k=1}^n y_k$ since $\alpha_1 > 0$ is arbitrary. Then denote by $\mathcal{P}$ the set of all subsets of $\{1, \ldots, n\}$ of size k and take $\alpha_1 = \cdots = \alpha_k > 1$, $\alpha_{k+1} = \cdots = \alpha_n = 1$. By our hypotheses,

$$\sum_{S \in \mathcal{P}} \alpha_1^{\sum_{k \in S} x_k} \leq \sum_{S \in \mathcal{P}} \alpha_1^{\sum_{k \in S} y_k}.$$

If $\sum_{j=1}^k x_j^{\downarrow} > \sum_{j=1}^k y_j^{\downarrow}$, this leads to a contradiction for α_1 large enough. Thus $x \prec y$.

(iv) $\Rightarrow$ (i) Assume that $P = (p_{jk})_{j,k=1}^n$. Since $x_k = \sum_j y_j p_{jk}$, where $\sum_j p_{jk} = 1$, it follows from the definition of convexity that

$$f(x_k) \leq \sum_j p_{jk} f(y_j).$$

Using the relation $\sum_k p_{jk} = 1$, we infer that

$$\sum_{k=1}^n f(x_k) \leq \sum_{k=1}^n \sum_{j=1}^n p_{jk} f(y_j) = \sum_{j=1}^n \sum_{k=1}^n p_{jk} f(y_j) = \sum_{j=1}^n f(y_j).$$

(v) $\Rightarrow$ (iv) Since T-transformations are doubly stochastic, the product of T-transformations is a doubly stochastic transformation.

(i) $\Rightarrow$ (v) Let x and y be two distinct points of $\mathbb{R}^n$ such that $x \prec y$. Since permutations are T-transformations, we may assume that their components verify the conditions

$$x_1 \geq x_2 \geq \cdots \geq x_n \quad \text{and} \quad y_1 \geq y_2 \geq \cdots \geq y_n.$$

Let j be the largest index such that $x_j < y_j$ and let k be the smallest index such that $k > j$ and $x_k > y_k$. The existence of such a pair of indices is motivated by the fact that the largest index i with $x_i \neq y_i$ verifies $x_i > y_i$. Then

$$y_j > x_j \geq x_k > y_k.$$

Put $\varepsilon = \min\{y_j - x_j, x_k - y_k\}$, $\lambda = 1 - \varepsilon/(y_j - y_k)$ and

$$y^* = (y_1, \ldots, y_{j-1}, y_j - \varepsilon, y_{j+1}, \ldots, y_{k-1}, y_k + \varepsilon, y_{k+1}, \ldots, y_n).$$

Clearly, $\lambda \in (0,1)$. Letting Q be the permutation matrix which interchanges the components of order j and k, we see that $y^* = Ty$ for the representation

$$T = \lambda I + (1 - \lambda)Q.$$

From (v) $\Rightarrow$ (iv) $\Rightarrow$ (i) it follows that $y^* \prec y$. On the other hand, $x \prec y^*$. In fact,

$$\begin{gathered}
\sum_{r=1}^{s} y_r^* = \sum_{r=1}^{s} y_r \geq \sum_{r=1}^{s} x_r \quad \text{for } s = 1, \ldots, j-1 \\
y_j^* \geq x_j \quad \text{and} \quad y_r^* = y_r \quad \text{for } r = j+1, \ldots, k-1 \\
\sum_{r=1}^{s} y_r^* = \sum_{r=1}^{s} y_r \geq \sum_{r=1}^{s} x_r \quad \text{for } s = k+1, \ldots, n \\
\sum_{r=1}^{n} y_r^* = \sum_{r=1}^{n} y_r = \sum_{r=1}^{n} x_r.
\end{gathered}$$

Letting $d(u,v)$ be the number of indices r such that $u_r \neq v_r$, it is clear that $d(x, y^*) \leq d(x,y) - 1$, so repeating the above algorithm (at most) $n-1$ times, we arrive at x. $\square$

The theory of majorization has important applications to statistics, combinatorics, metric geometry, and eigenvalue distribution of compact operators. Some of them are summarized in the classical book by A. W. Marshall and I. Olkin [155]. In the Comments section of Chapter 4 we shall present an important component of this theory, Schur convexity, to which I. Schur was led by the following result:

Theorem 1.10.2 (I. Schur [224]) *Let A be an $n \times n$-dimensional Hermitian matrix with diagonal elements $a_{11}, \ldots, a_{nn}$ and eigenvalues $\lambda_1, \ldots, \lambda_n$. Then*

$$(a_{11}, \ldots, a_{nn}) \prec (\lambda_1, \ldots, \lambda_n).$$

Proof. By the spectral decomposition theorem, $A = UDU^*$, where $U = (u_{kl})_{k,l}$ is unitary and D is diagonal, with entries $\lambda_1, \ldots, \lambda_n$. Then the diagonal elements of A are

$$a_{kk} = \sum_{l=1}^{n} u_{kl}\bar{u}_{kl}\lambda_l = \sum_{l=1}^{n} p_{kl}\lambda_l,$$

where $p_{kl} = u_{kl}\bar{u}_{kl}$. Since U is unitary, the matrix $P = (p_{kl})_{k,l}$ is doubly stochastic and Theorem 1.10.1 applies. $\square$

Corollary 1.10.3 (Hadamard's inequality) *If A is an $n \times n$-dimensional positive matrix with diagonal elements $a_{11}, \ldots, a_{nn}$ and eigenvalues $\lambda_1, \ldots, \lambda_n$, then*

$$\prod_{k=1}^{n} a_{kk} \geq \prod_{k=1}^{n} \lambda_k.$$

An alternative form of this corollary is as follows: *If $A = (a_{jk})_{j,k}$ is an $m \times n$-dimensional complex matrix, then*

$$\det AA^* \leq \prod_{j=1}^{m} \sum_{k=1}^{n} |a_{jk}|^2.$$

A. Horn [109] (see also [155]) proved a converse to Theorem 1.10.2. Namely, if x and y are two vectors in $\mathbb{R}^n$ such that $x \prec y$, then there exists a symmetric matrix A such that the entries of x are the diagonal elements of A and the entries of y are the eigenvalues of A. We are thus led to the following example of a *moment map*. Let α be an n-tuple of real numbers and let $\mathcal{O}_\alpha$ be the set of all symmetric matrices in $\mathrm{M}_n(\mathbb{R})$ with eigenvalues α. Consider the map $\Phi\colon \mathcal{O}_\alpha \to \mathbb{R}^n$ that takes a matrix to its diagonal. Then the image of Φ is a convex polyhedron, whose vertices are the $n!$ permutations of α. See M. Atiyah [12] for a large generalization and a surprising link between mechanics, Lie group theory and spectra of matrices.

We end this section with a result concerning a weaker relation of majorization (see Exercise 5, Section 2.7, for a generalization):

Theorem 1.10.4 (M. Tomić [236] and H. Weyl [245]) *Let $f\colon I \to \mathbb{R}$ be a nondecreasing convex function. If $(a_k)_{k=1}^n$ and $(b_k)_{k=1}^n$ are two families of numbers in I with $a_1 \geq \cdots \geq a_n$ and*

$$\sum_{k=1}^{m} a_k \leq \sum_{k=1}^{m} b_k \quad \textit{for } m = 1, \ldots, n$$

then

$$\sum_{k=1}^{n} f(a_k) \leq \sum_{k=1}^{n} f(b_k).$$

Proof. (By mathematical induction.) The case $n = 1$ is clear. Assume the conclusion works for all families of length $n - 1$. By our hypotheses, $a_1 \leq b_1$.

If $a_1 = b_1$ or $a_n \le b_n$, then the conclusion will follow from the induction hypothesis and the monotonicity of f. If $a_1 < b_1$ and $a_n > b_n$, then the points $a_1, \dots, a_n$ are interior to I and Lemma 1.5.1 yields a nondecreasing function $\varphi\colon \operatorname{int} I \to \mathbb{R}$ such that $\varphi(x) \in \partial f(x)$ for all x. Moreover, $\varphi \ge 0$ since f is nondecreasing. See Proposition 1.6.1. As in the proof of Theorem 1.5.4, we may conclude that

$$\begin{aligned}\sum_{k=1}^{n}(f(b_k) - f(a_k)) &\ge \sum_{k=1}^{n}\varphi(a_k)(b_k - a_k)\\ &= \varphi(a_n)\Big[\sum_{k=1}^{n} b_k - \sum_{k=1}^{n} a_k\Big]\\ &\qquad + \sum_{m=1}^{n-1}\Big[(\varphi(a_m) - \varphi(a_{m+1}))\sum_{k=1}^{m}(b_k - a_k)\Big]\\ &\ge 0\end{aligned}$$

and the proof is done. □

Exercises

1. Notice that $(1/n, 1/n, \dots, 1/n) \prec (1, 0, \dots, 0)$ and infer from Muirhead's inequality (the equivalence (i) $\Leftrightarrow$ (iii) in Theorem 1.10.1) the AM–GM inequality.

2. (I. Schur [222]) Consider the matrix $A = (a_{jk})_{j,k=1}^{n}$, whose eigenvalues are $\lambda_1, \dots, \lambda_n$. Prove that

$$\sum_{j,k=1}^{n} |a_{jk}|^2 \ge \sum_{k=1}^{n} |\lambda_k|^2.$$

3. Apply the result of the preceding exercise to derive the AM–GM inequality.
[*Hint*: Write down an $n\times n$ matrix whose nonzero entries are $x_1, \dots, x_n > 0$ and whose characteristic polynomial is $x^n - \prod_{k=1}^{n} x_k$.]

4. (The rearrangement inequalities of Hardy–Littlewood–Pólya [99]) Let $x_1, \dots, x_n, y_1, \dots, y_n$ be real numbers. Prove that

$$\sum_{k=1}^{n} x_k^{\downarrow} y_{n-k+1}^{\downarrow} \le \sum_{k=1}^{n} x_k y_k \le \sum_{k=1}^{n} x_k^{\downarrow} y_k^{\downarrow}.$$

If the numbers $x_1, \dots, x_n, y_1, \dots, y_n$ are nonnegative, prove that

$$\prod_{k=1}^{n}(x_k^{\downarrow} + y_k^{\downarrow}) \le \prod_{k=1}^{n}(x_k + y_k) \le \prod_{k=1}^{n}(x_k^{\downarrow} + y_{n-k+1}^{\downarrow}).$$

5. Give another proof of the AM–GM inequality by applying the rearrangement inequalities of Hardy–Littlewood–Pólya.

6. (An overview on symmetric-decreasing rearrangements, see [144] for a full account) Let (X, Σ, μ) be a measure space. The *distribution function* of a μ-measurable function $f\colon X \to \mathbb{C}$ is defined by

$$\lambda_f(\alpha) = \mu(\{x \bigm| |f(x)| > \alpha\}) \quad \text{for } \alpha > 0.$$

(i) Infer from Fubini's theorem that

$$\int_X |f(x)|^p \, d\mu = p \int_0^\infty \alpha^{p-1} \lambda_f(\alpha) \, d\alpha$$

for every $p > 0$ and every $f \in L^p(\mu)$. The particular case where $f \geq 0$, $\mu = \delta_x$ and $p = 1$ is known as the *layer cake representation* of f.

(ii) The *symmetric-decreasing rearrangement* of f is the function

$$f^{\downarrow}(\alpha) = \inf\{t \mid \lambda_f(t) \leq \alpha\}.$$

Consider a discrete measure space and conclude that the symmetric-decreasing rearrangement of a finite sequence $x_1, x_2, \dots, x_n$ of nonnegative numbers is $x_1^{\downarrow} \geq x_2^{\downarrow} \geq \cdots \geq x_n^{\downarrow}$.

(iii) (Equimeasurability of f and $f^{\downarrow}$) Suppose that $f \in L^p(\mu)$, where $(1 \leq p < \infty)$. Prove that

$$\chi^{\downarrow}_{\{x \mid |f(x)| > t\}} = \chi_{\{x \mid f^{\downarrow}(x) > t\}} \quad \text{for all } t > 0,$$

and conclude that $\|f\|_{L^p} = \|f^{\downarrow}\|_{L^p}$.

7. (An integral version of the Hardy–Littlewood–Pólya inequality) Let f and g be two integrable functions on $[0, 1]$, taking values in an interval I. Prove that

$$\int_0^1 \varphi(f(x)) \, dx \leq \int_0^1 \varphi(g(x)) \, dx$$

for every continuous convex function $\varphi\colon I \to \mathbb{R}$ (for which both functions $\varphi \circ f$ and $\varphi \circ g$ are integrable) if and only if $f \prec g$, that is,

$$\int_0^x f^{\downarrow}(t) \, dt \leq \int_0^x g^{\downarrow}(t) \, dt \quad \text{for } 0 \leq x < 1$$

and

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

1.11 Comments

The recognition of convex functions as a class of functions to be studied in its own right generally can be traced back to J. L. W. V. Jensen [115]. However, he was not the first one to deal with convex functions. The discrete form of Jensen's inequality was first proved by O. Hölder [106] in 1889, under the stronger hypothesis that the second derivative is nonnegative. Moreover, O. Stolz [233] proved in 1893 that every midpoint convex continuous function $f\colon [a,b] \to \mathbb{R}$ has left and right derivatives at each point of (a,b).

While the usual convex functions are continuous at all interior points (a fact due to J. L. W. V. Jensen [115]), the midpoint convex functions may be discontinuous everywhere. In fact, regard $\mathbb{R}$ as a vector space over $\mathbb{Q}$ and choose (via the axiom of choice) a basis $(b_i)_{i\in I}$ of $\mathbb{R}$ over $\mathbb{Q}$, that is, a maximal linearly independent set. Then every element x of $\mathbb{R}$ has a unique representation $x = \sum_{i\in I} c_i(x) b_i$ with coefficients $c_i(x)$ in $\mathbb{Q}$ and $c_i(x) = 0$ except for finitely many indices i. The uniqueness of this representation gives rise, for each $i \in I$, of a coordinate projection $\mathrm{pr}_i\colon x \to c_i(x)$, from $\mathbb{R}$ onto $\mathbb{Q}$. As G. Hamel [95] observed in 1905, the functions pr_i are discontinuous everywhere and

$$\mathrm{pr}_i(\alpha x + \beta y) = \alpha \, \mathrm{pr}_i(x) + \beta \, \mathrm{pr}_i(y),$$

for all $x, y \in \mathbb{R}$ and all $\alpha, \beta \in \mathbb{Q}$.

H. Blumberg [31] and W. Sierpiński [226] have noted independently that if $f\colon (a,b) \to \mathbb{R}$ is measurable and midpoint convex, then f is also continuous (and thus convex). See [212, pp. 220–221] for related results. The complete understanding of midpoint convex functions is due to G. Rodé [214], who proved that a real-valued function is midpoint convex if and only if it is the pointwise supremum of a family of functions of the form $a + c$, where a is additive and c is a real constant.

Popoviciu's inequality [206], as stated in Theorem 1.1.8, was known to him in a more general form, which applies to all continuous convex functions $f\colon I \to \mathbb{R}$ and all finite families $x_1, \dots, x_n$ of $n \geq 2$ points with equal weights. Later on, this fact was extended by P. M. Vasić and Lj. R. Stanković to the case of arbitrary weights $\lambda_1, \dots, \lambda_n > 0$:

$$\begin{aligned}
&\sum_{1\leq i_1 < \cdots < i_k \leq n} (\lambda_{i_1} + \cdots + \lambda_{i_k}) f\Big(\frac{\lambda_{i_1} x_{i_1} + \cdots + \lambda_{i_k} x_{i_k}}{\lambda_{i_1} + \cdots + \lambda_{i_k}}\Big) \\
&\qquad \leq \binom{n-2}{k-2} \Big[\frac{n-k}{k-1} \sum_{i=1}^{n} \lambda_i f(x_i) + \Big(\sum_{i=1}^{n} \lambda_i\Big) f\Big(\frac{\lambda_1 x_1 + \cdots + \lambda_n x_n}{\lambda_1 + \cdots + \lambda_n}\Big)\Big].
\end{aligned}$$

See [196, Section 6.1].

The Rogers–Hölder inequality (known to most mathematicians as the Hölder inequality) was proved in 1888 by L. J. Rogers [215] in a slightly different, but equivalent form. The basic ingredient was his weighted form of the

AM–GM inequality (as stated in Theorem 1.1.6). One year later, O. Hölder [106] clearly wrote that he, after Rogers, proved the inequality

$$\Big(\sum_{k=1}^{n} a_k b_k\Big)^t \leq \Big(\sum_{k=1}^{n} a_k\Big)^{t-1} \Big(\sum_{k=1}^{n} a_k b_k^t\Big),$$

valid for all $t > 1$, and all $a_k > 0$, $b_k > 0$, $k = 1, \ldots, n$, $n \in \mathbb{N}^*$. His idea was to apply Jensen's inequality to the function $f(x) = x^t$, $x > 0$. However, F. Riesz was the first who stated and used the Rogers–Hölder inequality as we did in Section 1.2. See the paper of L. Maligranda [152] for the complete history.

The inequality of Jakob Bernoulli, $(1+a)^n \geq 1+na$, for all $a \geq -1$ and n in $\mathbb{N}$, appeared in [22]. The generalized form (see Exercise 2, Section 1.1) is due to O. Stolz and J. A. Gmeiner; see [152]. The classical AM–GM inequality can be traced back to C. Maclaurin [149]; see [99, p. 52]. L. Maligranda [152] noticed that the classical Bernoulli inequality, the classical AM–GM inequality and the generalized AM–GM inequality of L. J. Rogers are all equivalent (that is, each one can be used to prove the other ones).

The upper estimate for Jensen's inequality given in Theorem 1.4.1 is due to C. P. Niculescu [179].

A refinement of Jensen's inequality for "more convex" functions was proved by S. Abramovich, G. Jameson and G. Sinnamon [1]. Call a function $\varphi\colon [0,\infty) \to \mathbb{R}$ *superquadratic* provided that for each $x \geq 0$ there exists a constant $C_x \in \mathbb{R}$ such that $\varphi(y) - \varphi(x) - \varphi(|y-x|) \geq C_x(y-x)$ for all $y \geq 0$. For example, if $\varphi\colon [0,\infty) \to \mathbb{R}$ is continuously differentiable, $\varphi(0) \leq 0$ and either $-\varphi'$ is subadditive or $\varphi'(x)/x$ is nondecreasing, then φ is superquadratic. Particularly, this is the case of $x^2 \log x$. Moreover every superquadratic nonnegative function is convex. Their main result asserts that the inequality

$$\varphi\Big(\int_X f(y)\,d\mu(y)\Big) \leq \int_X \Big[\varphi(f(x)) - \varphi\Big(\Big|f(x) - \int_X f(y)\,d\mu(y)\Big|\Big)\Big]\,d\mu(x)$$

holds for all probability spaces (X, Σ, μ) and all nonnegative μ-measurable functions f if and only if φ is superquadratic.

The proof of Theorem 1.5.6 (the Jensen–Steffensen inequality) is due to J. Pečarić [195].

The history of Hardy's inequality is told in Section 9.8 of [99]. Its many ramifications and beautiful applications are the subject of two monographs, [135] and [192].

We can arrive at Hardy's inequality via mixed means. For a positive n-tuple $a = (a_1, \ldots, a_n)$, the *mixed arithmetic-geometric inequality* asserts that the arithmetic mean of the numbers

$$a_1, \sqrt{a_1 a_2}, \ldots, \sqrt[n]{a_1 a_2 \cdots a_n}$$

does not exceed the geometric mean of the numbers

$$a_1, \frac{a_1+a_2}{2}, \ldots, \frac{a_1+a_2+\cdots+a_n}{n}$$

(see K. Kedlaya [123]). As noted by B. Mond and J. Pečarić [170], the arithmetic and the geometric means can be replaced (in this order) by any pair (M_r, M_s) of power means with $r > s$. For $r = p > 1$ and $s = 1$ this gives us

$$\Big[\frac{1}{n}\sum_{k=1}^{n}\Big(\frac{a_1+a_2+\cdots+a_k}{k}\Big)^p\Big]^{1/p} \le \frac{1}{n}\sum_{k=1}^{n}\Big(\frac{1}{k}\sum_{j=1}^{k} a_j^p\Big)^{1/p}$$

so that $\sum_{k=1}^{n}((a_1+a_2+\cdots+a_k)/k)^p$ is less than or equal to

$$n^{1-p}\Big(\sum_{j=1}^{n} a_j^p\Big)\Big[\sum_{k=1}^{n}\Big(\frac{1}{k}\Big)^{1/p}\Big]^p \le \Big(\frac{p}{p-1}\Big)^p\Big(\sum_{j=1}^{n} a_j^p\Big),$$

as $\int_0^n x^{-1/p}\,dx = \frac{p}{p-1}n^{1-1/p}$. The integral case of this approach is discussed by A. Čižmešija and J. Pečarić [54].

Carleman's inequality (and its ramifications) has also received a great deal of attention in recent years. The reader may consult the papers by J. Pečarić and K. Stolarsky [198], J. Duncan and C. M. McGregor [68], M. Johansson, L.-E. Persson and A. Wedestig [116], S. Kaijser, L.-E. Persson and A. Öberg [118], A. Čižmešija and J. Pečarić and L.-E. Persson [55].

The complete forms of Jensen's and Chebyshev's inequalities (Theorems 1.8.4 and 1.8.6 above) are due to C. P. Niculescu [179]. The smooth variant of Corollary 1.8.5 was first noticed by S. S. Dragomir and N. M. Ionescu [67]. An account on the history, variations and generalizations of the Chebyshev inequality can be found in the paper by D. S. Mitrinović and P. M. Vasić [169]. Other complements to Jensen's and Chebyshev's inequalities can be found in the papers by H. Heinig and L. Maligranda [100] and S. M. Malamud [150].

The dramatic story of the Hermite–Hadamard inequality is told in a short note by D. S. Mitrinović and I. B. Lacković [167]: In a letter sent on November 22, 1881, to *Mathesis* (and published there in 1883), Ch. Hermite [102] noted that every convex function $f\colon [a,b] \to \mathbb{R}$ satisfies the inequalities

$$f\Big(\frac{a+b}{2}\Big) \le \frac{1}{b-a}\int_a^b f(x)\,dx \le \frac{f(a)+f(b)}{2}$$

and illustrated this with Example 1.9.1 (i) in our text. Ten years later, the left-hand side inequality was rediscovered by J. Hadamard [94]. However the priority of Hermite was not recorded and his note was not even mentioned in Hermite's *Collected Papers* (published by E. Picard).

The precision in the Hermite–Hadamard inequality can be estimated via two classical inequalities that work in the Lipschitz function framework. Suppose that $f\colon [a,b] \to \mathbb{R}$ is a Lipschitz function, with *Lipschitz constant*

$$\mathrm{Lip}(f) = \sup\Big\{\Big|\frac{f(x)-f(y)}{x-y}\Big| \mid x \neq y\Big\}.$$

Then the left Hermite–Hadamard inequality can be estimated by the *inequality of Ostrowski*,

$$\Big|f(x) - \frac{1}{b-a}\int_a^b f(t)\,dt\Big| \leq M\Big[\frac{1}{4} + \Big(\frac{x-\frac{a+b}{2}}{b-a}\Big)^2\Big](b-a),$$

while the right Hermite–Hadamard inequality can be estimated by the *inequality of Iyengar*,

$$\Big|\frac{f(a)+f(b)}{2} - \frac{1}{b-a}\int_a^b f(t)\,dt\Big| \leq \frac{M(b-a)}{4} - \frac{1}{4M(b-a)}\,(f(b)-f(a))^2,$$

where $M = \mathrm{Lip}(f)$. The first inequality is a direct consequence of the triangle inequality. The second one will be proved in Section 4.3.

A dual Hermite–Hadamard inequality is discussed by C. P. Niculescu [180]. A complete extension of the Hermite–Hadamard inequality to the class of n-convex functions was recently obtained by M. Bessenyei and Z. Páles [25]. For other results on the Hermite–Hadamard inequality, see the monograph [196].

The theory of majorization (and its applications) is presented in great detail in the book by A. W. Marshall and I. Olkin [155]. The integral version of majorization is based on the concept of rearrangement. A modern exposition of the theory can be found in E. H. Lieb and M. Loss [144] (but [99] is still a big source of concrete examples).

As noticed by L. Maligranda [151], if f is a nonnegative convex (concave) function on an interval $I = [0, a]$, then its distribution function λ_f is convex on $[\operatorname{ess\,inf}_I f, \infty)$ (concave on $[0, \operatorname{ess\,sup}_I f)$) and $f^{\downarrow}$ is convex on $[0,\infty)$. As an application he derived the following two inequalities:

Favard's inequality [76]. *Let f be a nonnegative continuous concave function on $[0, a]$, not identically zero, and let φ be a convex function on $[0, 2M_1(f)]$. Then*

$$\frac{1}{a}\int_0^a \varphi(f(x))\,dx \leq \frac{1}{2M_1(f)}\int_0^{2M_1(f)} \varphi(t)\,dt.$$

Berwald's inequality [24]. *Let f be a nonnegative continuous concave function on $[0, a]$ and $0 < r \leq p$. Then*

$$\Big(\frac{1}{a}\int_0^a f^p(x)\,dx\Big)^{1/p} \leq \frac{(1+r)^{1/r}}{(1+p)^{1/p}}\Big(\frac{1}{a}\int_0^a f^r(x)\,dx\Big)^{1/r}.$$

We end this chapter with a brief discussion on the differentiability properties of convex functions from a generic point of view.

Let $\mathcal{P}$ be a property which refers to the elements of a complete metric space X. We say that $\mathcal{P}$ is *generic* (or that most elements of X enjoy $\mathcal{P}$) if those elements not enjoying the property $\mathcal{P}$ form a set of first Baire category, that is, a countable union of nowhere dense sets. The space $C[0,1]$, of all continuous real functions on $[0,1]$, endowed with the usual sup-norm, is complete. The same is true for $\operatorname{Conv}[0,1]$, the subset of all continuous convex functions on $[0,1]$. A well-known elegant proof of S. Banach shows that most functions in $C[0,1]$ are nowhere differentiable. The situation in $\operatorname{Conv}[0,1]$ is different. In fact, as noted by V. Klee [128], *most convex functions in* $C[0,1]$ *are differentiable.*

The generic aspects of the second differentiability of convex functions are described by T. Zamfirescu [249]: *For most convex functions* $f\colon [0,1] \to \mathbb{R}$*,*

$$\underline{D}f' = 0 \quad \text{or} \quad \overline{D}f' = \infty \quad \text{everywhere.}$$

Moreover, for most convex functions f*, the second derivative* f'' *vanishes wherever it exists, that is, almost everywhere.* Thus, the behavior of the integral of Riesz's increasing function φ mentioned in Remark 1.6.2 is rather the rule, not the exception.

Most convex functions have, however, many points where f'' does not exist. More precisely, *for most convex functions* f*, at most points of* $[0,1]$*,* $\underline{D}f' = 0$ *and* $\overline{D}f' = \infty$. See T. Zamfirescu [250]. This is complemented by another result (also by T. Zamfirescu [251]), showing that *for most convex functions* f*, at densely, uncountably many points in* $[0,1]$*,* $f'' = \infty$*.*

2

Comparative Convexity on Intervals

This chapter is devoted to a succinct presentation of several classes of functions acting on intervals, which satisfy inequalities of the form

$$f(M(x,y)) \leq N(f(x), f(y)),$$

for a suitable pair of means M and N. Leaving out the case of usual convex functions (when M and N coincide with the arithmetic mean), the most important classes that arise in applications are:

- the class of log-convex functions (M is the arithmetic mean and N is the geometric mean)
- the class of multiplicatively convex functions (M and N are both geometric means)
- the class of M_p-convex functions (M is the arithmetic mean and N is the power mean of order p).

They all provide important applications to many areas of mathematics.

2.1 Algebraic Versions of Convexity

The usual definition of a convex function (of one real variable) depends on the structure of $\mathbb{R}$ as an ordered vector space. As $\mathbb{R}$ is actually an ordered field, it is natural to investigate what happens when addition is replaced by multiplication and the arithmetic mean is replaced by the geometric mean.

The characteristic property of the subintervals I of $\mathbb{R}$ is

$$x, y \in I \text{ and } \lambda \in [0,1] \implies (1-\lambda)x + \lambda y \in I$$

so, in order to draw a parallel in the multiplicative case, we must restrict to the subintervals J of $(0,\infty)$ and use instead the following fact:

$$x, y \in J \text{ and } \lambda \in [0,1] \implies x^{1-\lambda} y^{\lambda} \in J.$$

Depending on which type of mean, arithmetic (A) or geometric (G), we consider on the domain and on the range, we shall encounter one of the following four classes of functions:

- (A, A)-*convex functions*, the usual convex functions;
- (A, G)-*convex functions*;
- (G, A)-*convex functions*;
- (G, G)-*convex functions*.

More precisely, the (A, G)-*convex functions* (usually known as log-*convex functions*) are those functions $f\colon I \to (0, \infty)$ for which

$$x, y \in I \text{ and } \lambda \in [0, 1] \implies f((1-\lambda)x + \lambda y) \leq f(x)^{1-\lambda} f(y)^{\lambda}, \tag{AG}$$

that is, for which $\log f$ is convex. If a function $f\colon I \to \mathbb{R}$ is log-convex, then it is also convex. In fact, according to the AM–GM inequality,

$$f((1-\lambda)x + \lambda y) \leq f(x)^{1-\lambda} f(y)^{\lambda} \leq (1-\lambda) f(x) + \lambda f(y).$$

The converse does not work. For example, the function $\mathrm{e}^x - 1$ is convex and log-concave.

One of the most notable examples of a log-convex function is Euler's *gamma function*,

$$\Gamma(x) = \int_0^\infty t^{x-1} \mathrm{e}^{-t}\, dt, \quad x > 0.$$

The place of Γ in the landscape of log-convex functions is the subject of the next section.

The class of all (G, A)-*convex functions* consists of all real-valued functions f (defined on subintervals I of $(0, \infty)$) for which

$$x, y \in I \text{ and } \lambda \in [0, 1] \implies f(x^{1-\lambda} y^{\lambda}) \leq (1-\lambda) f(x) + \lambda f(y). \tag{GA}$$

In the context of twice-differentiable functions $f\colon I \to \mathbb{R}$, (G, A)-convexity means $x^2 f'' + x f' \geq 0$.

The (G, G)-*convex functions* (called *multiplicatively convex functions* in what follows) are those functions $f\colon I \to J$ (acting on subintervals of $(0, \infty)$) such that

$$x, y \in I \text{ and } \lambda \in [0, 1] \implies f(x^{1-\lambda} y^{\lambda}) \leq f(x)^{1-\lambda} f(y)^{\lambda}. \tag{GG}$$

Equivalently, f is multiplicatively convex if and only if $\log f(x)$ is a convex function of $\log x$. This fact will be shown in Lemma 2.3.1 below. Due to the arithmetic-geometric mean inequality, all multiplicatively convex functions (and also all nondecreasing convex functions) are (G, A)-convex functions.

The theory of multiplicatively convex functions is similar to that of classical convex functions. In fact, they differ from each other only by a change of variable and a change of function:

Lemma 2.1.1 *Suppose that I is a subinterval of $(0,\infty)$ and $f\colon I \to (0,\infty)$ is a multiplicatively convex function on I. Then*

$$F = \log\circ f\circ\exp\colon\ \log(I) \to \mathbb{R}$$

is a convex function. Conversely, if J is an interval and $F\colon J \to \mathbb{R}$ is a convex function, then

$$f = \exp\circ F\circ\log\colon\ \exp(J) \to (0,\infty)$$

is a multiplicatively convex function.

The proof is straightforward. Lemma 2.1.1 can be adapted easily to other situations and allows us to deduce new inequalities from old ones. This idea is central to Section 2.3 below.

Exercises

1. (Some geometrical consequences of log-convexity)

(i) A convex quadrilateral $ABCD$ is inscribed in the unit circle. Its sides satisfy the inequality $AB \cdot BC \cdot CD \cdot DA \geq 4$. Prove that $ABCD$ is a square.

(ii) Suppose that A, B, C are the angles of a triangle, expressed in radians. Prove that

$$\sin A \sin B \sin C < \left(\frac{3\sqrt{3}}{2\pi}\right)^3 ABC < \left(\frac{\sqrt{3}}{2}\right)^3,$$

unless $A = B = C$.

[*Hint*: Note that the sine function is log-concave, while $x/\sin x$ is log-convex on $(0,\pi)$.]

2. Let (X, Σ, μ) be a measure space and let $f\colon X \to \mathbb{C}$ be a measurable function, which is in $L^t(\mu)$ for t in a subinterval I of $(0,\infty)$. Infer from the Cauchy–Buniakovski–Schwarz inequality that the function $t \to \log\int_X |f|^t\,d\mu$ is convex on I.

Remark. The result of this exercise is equivalent to *Lyapunov's inequality* [148]: *If $a \geq b \geq c$, then*

$$\left(\int_X |f|^b\,d\mu\right)^{a-c} \leq \left(\int_X |f|^c\,d\mu\right)^{a-b}\left(\int_X |f|^a\,d\mu\right)^{b-c}$$

(provided the integrability aspects are fixed). Equality holds if and only if one of the following conditions hold:

(i) f is constant on some subset of Ω and 0 elsewhere;

(ii) $a = b$;

(iii) $b = c$;

(iv) $c(2a-b)=ab$.

3. (P. Montel [171]) Let I be an interval. Prove that the following assertions are equivalent for every function $f\colon I \to (0,\infty)$:

(i) f is log-convex;
(ii) the function $x \to \mathrm{e}^{\alpha x} f(x)$ is convex on I for all $\alpha \in \mathbb{R}$;
(iii) the function $x \to [f(x)]^{\alpha}$ is convex on I for all $\alpha > 0$.

[*Hint*: For (iii) $\Rightarrow$ (i), note that $([f(x)]^{\alpha}-1)/\alpha$ is convex for all $\alpha > 0$ and $\log f(x) = \lim_{\alpha\to 0+}([f(x)]^{\alpha}-1)/\alpha$. Then apply Corollary 1.3.8.]

4. Prove that the sum of two log-convex functions is also log-convex. [*Hint*: Note that this assertion is equivalent to the following inequality for positive numbers: $a^{\alpha}b^{\beta} + c^{\alpha}d^{\beta} \le (a+c)^{\alpha}(b+d)^{\beta}$.]

5. (S. Simic [227]) Let $(a_n)_n$ be a sequence of positive numbers. Prove that the following assertions are equivalent:

(i) $(a_n)_n$ is log-convex (that is, $a_{n-1}a_{n+1} \ge a_n^2$ for all $n \ge 1$);
(ii) for each $x \ge 0$, the sequence $P_n(x) = \sum_{k=0}^{n} a_k \binom{n}{k} x^{n-k}$ $(n \in \mathbb{N})$ is log-convex.

6. A function $f\colon (0,\infty) \to \mathbb{R}$ is called *completely monotonic* if f has derivatives of all orders and satisfies $(-1)^n f^{(n)}(x) \ge 0$ for all $x > 0$ and $n \in \mathbb{N}$. In particular, completely monotonic functions are decreasing and convex.

(i) Prove that

$$(-1)^{nk}(f^{(k)}(x))^n \le (-1)^{nk}(f^{(n)}(x))^k (f(x))^{n-k}$$

for all $x > 0$ and all integers n, k with $n \ge k \ge 0$. Infer that any completely monotonic function is actually log-convex.

(ii) Prove that the function

$$V_q(x) = \frac{\exp(x^2)}{\Gamma(q+1)} \int_x^{\infty} \mathrm{e}^{-t^2}(t^2-x^2)^q \, dt$$

is completely monotonic on $(0,\infty)$ if $q \in (-1,0]$.

2.2 The Gamma and Beta Functions

The gamma function $\Gamma\colon (0,\infty) \to \mathbb{R}$ is defined by the relation

$$\Gamma(x) = \int_0^{\infty} t^{x-1}\mathrm{e}^{-t}\, dt \quad \text{for } x > 0.$$

Theorem 2.2.1 *The gamma function has the following properties:*

(i) $\Gamma(x+1) = x\Gamma(x)$ *for all* $x > 0$;
(ii) $\Gamma(1) = 1$;

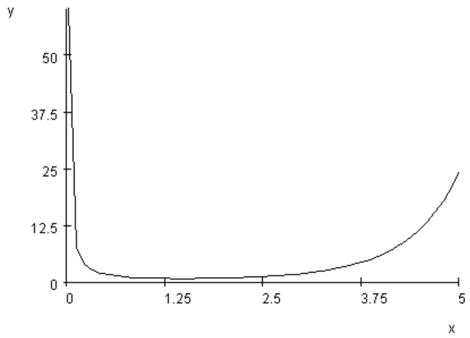

Fig. 2.1. The graph of Γ.

(iii) Γ *is* log-*convex.*

Proof. (i) Using integration by parts we get

$$\Gamma(x+1) = \int_0^\infty t^x \mathrm{e}^{-t}\,dt = [-t^x \mathrm{e}^{-t}]\Big|_{t=0}^{\infty} + x\int_0^\infty t^{x-1}\mathrm{e}^{-t}\,dt = x\Gamma(x)$$

for all $x > 0$.

The property (ii) is obvious.

(iii) Let $x, y > 0$ and let $\lambda, \mu \geq 0$ with $\lambda + \mu = 1$. Then, by the Rogers–Hölder inequality, we have

$$\begin{aligned}\Gamma(\lambda x + \mu y) &= \int_0^\infty t^{\lambda x + \mu y - 1}\mathrm{e}^{-t}\,dt = \int_0^\infty (t^{x-1}\mathrm{e}^{-t})^\lambda (t^{y-1}\mathrm{e}^{-t})^\mu\,dt \\ &\leq \left(\int_0^\infty t^{x-1}\mathrm{e}^{-t}\,dt\right)^\lambda \left(\int_0^\infty t^{y-1}\mathrm{e}^{-t}\,dt\right)^\mu = \Gamma^\lambda(x)\Gamma^\mu(y)\end{aligned}$$

which proves that Γ is log-convex. $\square$

Corollary 2.2.2 $\Gamma(n+1) = n!$ *for all* $n \in \mathbb{N}$.

Corollary 2.2.3 *The gamma function is convex and* $x\Gamma(x)$ *approaches* 1 *as* $x \to 0+$.

C. F. Gauss first noted that Γ attains its minimum at $x = 1.461632145\ldots$.

The gamma function is the unique log-convex extension of the factorial function:

Theorem 2.2.4 (H. Bohr and J. Mollerup [32], [10]) *Suppose the function* $f\colon (0,\infty) \to \mathbb{R}$ *satisfies the following three conditions:*

(i) $f(x+1) = xf(x)$ *for all* $x > 0$*;*
(ii) $f(1) = 1$*;*
(iii) f *is* log-*convex.*

Then $f = \Gamma$.

Proof. By induction, from (i) and (ii) we infer that $f(n+1) = n!$ for all $n \in \mathbb{N}$.

Now, let $x \in (0,1]$ and $n \in \mathbb{N}^\star$. Then by (iii) and (i),

$$\begin{aligned} f(n+1+x) &= f((1-x)(n+1) + x(n+2)) \\ &\leq [f(n+1)]^{1-x} \cdot [f(n+2)]^x \\ &= [f(n+1)]^{1-x} \cdot (n+1)^x \cdot [f(n+1)]^x \\ &= (n+1)^x \cdot f(n+1) \\ &= (n+1)^x \cdot n! \end{aligned}$$

and

$$\begin{aligned} n! = f(n+1) &= f(x(n+x) + (1-x)(n+1+x)) \\ &\leq [f(n+x)]^x \cdot [f(n+1+x)]^{1-x} \\ &= (n+x)^{-x} \cdot [f(n+1+x)]^x \cdot [f(n+1+x)]^{1-x} \\ &= (n+x)^{-x} \cdot f(n+1+x). \end{aligned}$$

Thus, since $f(n+1+x) = (n+x)(n-1+x)\cdots xf(x)$, we obtain

$$\left(1+\frac{x}{n}\right)^x \leq \frac{(n+x)(n-1+x)\cdots xf(x)}{n!\, n^x} \leq \left(1+\frac{1}{n}\right)^x,$$

which yields

$$f(x) = \lim_{n\to\infty} \frac{n!\, n^x}{(n+x)(n-1+x)\cdots x} \quad \text{for } x \in (0,1].$$

We shall show that the above formula is valid for all $x > 0$ so that f is uniquely determined by the conditions (i), (ii) and (iii). Since Γ satisfies all these three conditions, we must have $f = \Gamma$.

To end the proof, suppose that $x > 0$ and choose an integer number m such that $0 < x - m \leq 1$. According to (i) and what we have just proved, we get

$$\begin{aligned} f(x) &= (x-1)\cdots(x-m)f(x-m) \\ &= (x-1)\cdots(x-m)\cdot \lim_{n\to\infty} \frac{n!\, n^{x-m}}{(n+x-m)(n-1+x-m)\cdots(x-m)} \\ &= \lim_{n\to\infty} \Bigg(\frac{n!\, n^x}{(n+x)(n-1+x)\cdots x} \\ &\qquad\qquad \cdot \frac{(n+x)(n+x-1)\cdots(n+x-(m-1))}{n^m} \Bigg) \end{aligned}$$

$$= \lim_{n\to\infty} \frac{n!\,n^x}{(n+x)(n-1+x)\cdots x}$$
$$\cdot \lim_{n\to\infty}\left(\left(1+\frac{x}{n}\right)\left(1+\frac{x-1}{n}\right)\cdots\left(1+\frac{x-m+1}{n}\right)\right)$$
$$= \lim_{n\to\infty} \frac{n!\,n^x}{(n+x)(n-1+x)\cdots x}.$$

$\square$

Corollary 2.2.5 $\Gamma(x) = \lim_{n\to\infty} \frac{n!\,n^x}{(n+x)(n-1+x)\cdots x}$ *for all* $x > 0$.

Before establishing a fundamental identity linking the gamma and sine functions, we need to express $\sin x$ as an infinite product:

Theorem 2.2.6 (L. Euler) *For all real numbers* x,

$$\sin x = x\prod_{k=1}^{\infty}\left(1-\frac{x^2}{k^2\pi^2}\right).$$

Proof. De Moivre's formula shows that $\sin(2n+1)\theta$ is a polynomial of degree $2n+1$ in $\sin\theta$ (for each $n \in \mathbb{N}$, arbitrarily fixed). This polynomial has roots $\pm\sin(k\pi/(2n+1))$ for $k = 0, \ldots, n$. It follows that

$$\sin(2n+1)\theta = (2n+1)\sin\theta\prod_{k=1}^{n}\left(1-\frac{\sin^2\theta}{\sin^2\frac{k\pi}{2n+1}}\right).$$

Suppose that $x > 0$ and fix arbitrarily two integers m and n such that $x < m < n$. The last identity shows that

$$\frac{\sin x}{(2n+1)\sin\frac{x}{2n+1}} = \prod_{k=1}^{n}\left(1-\frac{\sin^2\frac{x}{2n+1}}{\sin^2\frac{k\pi}{2n+1}}\right).$$

Denote by a_k the k-th factor in this last product. Since $2\theta/\pi < \sin\theta < \theta$ when $0 < \theta < \pi/2$, we find that

$$0 < 1-\frac{x^2}{4k^2} < a_k < 1 \quad \text{for } m < k \le n,$$

which yields

$$1 > a_{m+1}\cdots a_n > \prod_{k=1}^{n}\left(1-\frac{x^2}{4k^2}\right) > 1-\frac{x^2}{4}\sum_{k=m+1}^{n}\frac{1}{k^2} > 1-\frac{x^2}{4m}.$$

Hence

$$\frac{\sin x}{(2n+1)\sin\frac{x}{2n+1}}$$

lies between

$$\left(1-\frac{x^2}{4m}\right)\prod_{k=1}^{n}\left(1-\frac{\sin^2\frac{x}{2n+1}}{\sin^2\frac{k\pi}{2n+1}}\right) \quad \text{and} \quad \prod_{k=1}^{n}\left(1-\frac{\sin^2\frac{x}{2n+1}}{\sin^2\frac{k\pi}{2n+1}}\right)$$

and so, letting $n \to \infty$, we deduce that $\sin x/x$ lies between

$$\left(1-\frac{x^2}{4m}\right)\prod_{k=1}^{\infty}\left(1-\frac{x^2}{k^2\pi^2}\right) \quad \text{and} \quad \prod_{k=1}^{\infty}\left(1-\frac{x^2}{k^2\pi^2}\right).$$

The proof ends by letting $m \to \infty$. $\square$

Theorem 2.2.7 *For all real x with $0 < x < 1$,*

$$\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin \pi x}.$$

Proof. In fact, by Corollary 2.2.5 and Theorem 2.2.6 above we infer that

$$\begin{aligned}\Gamma(x)\Gamma(1-x) &= \lim_{n\to\infty}\frac{n!\,n^x n!\,n^{1-x}}{(n+x)\cdots x\,(n+1-x)\cdots(1-x)}\\ &= \frac{1}{x\prod_{k=1}^{\infty}(1-x^2/k^2)} = \frac{\pi}{\sin \pi x}.\end{aligned}$$

$\square$

Corollary 2.2.8 $\Gamma(1/2) = \sqrt{\pi}$.

A variant of the last corollary is the formula

$$\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \mathrm{e}^{-t^2/2}\,dt = 1$$

which appears in many places in mathematics, statistics and natural sciences. Another beautiful consequence of Theorem 2.2.4 is the following:

Theorem 2.2.9 (The Gauss–Legendre duplication formula)

$$\Gamma\left(\frac{x}{2}\right)\Gamma\left(\frac{x+1}{2}\right) = \frac{\sqrt{\pi}}{2^{x-1}}\Gamma(x) \quad \textit{for all } x > 0.$$

Proof. Notice that the function

$$f(x) = \frac{2^{x-1}}{\sqrt{\pi}}\Gamma\left(\frac{x}{2}\right)\Gamma\left(\frac{x+1}{2}\right) \quad x > 0,$$

verifies the conditions (i)–(iii) in Theorem 2.2.4 and thus equals Γ. $\square$

We will prove Stirling's formula, which is an important tool in analytic number theory. We shall need the following lemma:

Lemma 2.2.10 *The sequence $(a_n)_n$, whose n-th term is*

$$a_n = \log n! - \left(n + \frac{1}{2}\right) \log n + n,$$

converges.

Proof. We shall show that the sequence is decreasing and bounded below. In fact,

$$a_n - a_{n+1} = \left(n + \frac{1}{2}\right) \log\left(1 + \frac{1}{n}\right) - 1 \geq 0$$

since by the Hermite–Hadamard inequality applied to the convex function $1/x$ on $[n, n+1]$ we have

$$\log\left(1 + \frac{1}{n}\right) = \int_n^{n+1} \frac{dx}{x} \geq \frac{1}{n + 1/2}.$$

A similar argument (applied to the concave function $\log x$ on $[u, v]$) yields

$$\int_u^v \log x \, dx \leq (v - u) \log \frac{u + v}{2},$$

so that (taking into account the monotonicity of the log function) we get

$$\begin{aligned}\int_1^n \log x \, dx &= \int_1^{1+1/2} \log x \, dx + \int_{1+1/2}^{2+1/2} \log x \, dx + \cdots + \int_{n-1/2}^n \log x \, dx \\ &\leq \frac{1}{2} \log \frac{3}{2} + \log 2 + \cdots + \log(n-1) + \frac{1}{2} \log n \\ &< \frac{1}{2} + \log n! - \frac{1}{2} \log n.\end{aligned}$$

Since

$$\int_1^n \log x \, dx = n \log n - n + 1,$$

we conclude that

$$a_n = \log n! - \left(n + \frac{1}{2}\right) \log n + n > \frac{1}{2}.$$

The result now follows. □

Theorem 2.2.11 (Stirling's formula) $n! \sim \sqrt{2\pi}\, n^{n+1/2} \mathrm{e}^{-n}$.

Proof. Under the notation of the previous lemma, put

$$b_n = \mathrm{e}^{a_n} = \frac{n!}{n^{n+1/2} \mathrm{e}^{-n}} \quad \text{for } n = 1, 2, \ldots .$$

Then the sequence $(b_n)_n$ converges to some $b > 0$. Thus

$$\frac{b_n^2}{b_{2n}} = \frac{2^{2n+1/2}(n!)^2}{n^{1/2}(2n)!} \to \frac{b^2}{b} = b \quad \text{as } n \to \infty.$$

For $n = 1, 2, \ldots$, let $c_n = \dfrac{n!\, n^{1/2}}{(n+\frac{1}{2}) \cdots \frac{3}{2} \cdot \frac{1}{2}}$. Then by Corollary 2.2.5, $(c_n)_n$ converges to $\Gamma(1/2) = \sqrt{\pi}$ as $n \to \infty$. Hence

$$\frac{b_n^2}{b_{2n}} = c_n \Big(1 + \frac{1}{2n}\Big)\sqrt{2} \to \sqrt{2\pi} \quad \text{as } n \to \infty,$$

which yields $b = \sqrt{2\pi}$. Consequently,

$$b_n = \frac{n!}{n^{n+1/2}\mathrm{e}^{-n}} \to \sqrt{2\pi} \quad \text{as} \quad n \to \infty$$

and the proof is now complete. □

Closely related to the gamma function is the *beta function* B, which is the real function of two variables defined by the formula

$$B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}\, dt \quad \text{for } x, y > 0.$$

Theorem 2.2.12 *The beta function has the following properties:*
(i) $B(x, y) = B(y, x)$ *and* $B(x+1, y) = \frac{x}{x+y} B(x, y)$;
(ii) $B(x, y)$ *is a* log-*convex function of* x *for each fixed* y;
(iii) $B(x, y) = \dfrac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$.

Proof. (i) The first formula is clear. For the second,

$$\begin{aligned} B(x+1, y) &= \int_0^1 t^x (1-t)^{y-1}\, dt \\ &= \int_0^1 (1-t)^{x+y-1} \Big(\frac{t}{1-t}\Big)^x dt \\ &= \Big[\frac{-(1-t)^{x+y}}{x+y} \Big(\frac{t}{1-t}\Big)^x\Big]_{t=0}^{t=1} + \int_0^1 \frac{x}{x+y}\, t^{x-1}(1-t)^{y-1}\, dt \\ &= \frac{x}{x+y}\, B(x, y). \end{aligned}$$

(ii) Let $a, b, y > 0$ and let $\lambda, \mu \geq 0$ with $\lambda + \mu = 1$. By the Rogers–Hölder inequality,

$$\begin{aligned} B(\lambda a + \mu b, y) &= \int_0^1 (t^{a-1}(1-t)^{y-1})^\lambda (t^{b-1}(1-t)^{y-1})^\mu \, dt \\ &\le \left(\int_0^1 t^{a-1}(1-t)^{y-1} \, dt \right)^\lambda \left(\int_0^1 t^{b-1}(1-t)^{y-1} \, dt \right)^\mu \\ &= B^\lambda(a, y) \cdot B^\mu(a, y). \end{aligned}$$

(iii) Let $y > 0$ be arbitrarily fixed and consider the function

$$\varphi_y(x) = \frac{\Gamma(x+y)B(x,y)}{\Gamma(y)}, \quad x > 0.$$

Then φ_y is a product of log-convex functions and so it is itself log-convex. Also,

$$\begin{aligned} \varphi_y(x+1) &= \frac{\Gamma(x+y+1)B(x+1,y)}{\Gamma(y)} \\ &= \frac{[(x+y)\Gamma(x+y)][x/(x+y)]B(x,y)}{\Gamma(y)} = x\varphi_y(x) \end{aligned}$$

for all $x > 0$ and

$$\varphi_y(1) = \frac{\Gamma(1+y)B(1,y)}{\Gamma(y)} = y \int_0^1 (1-t)^{y-1} \, dt = 1.$$

Thus $\varphi_y = \Gamma$ by Theorem 2.2.4, and the assertion (iii) is now clear. $\square$

Exercises

1. Prove that $\Gamma(n + \frac{1}{2}) = \frac{(2n)!\,\sqrt{\pi}}{n!\,4^n}$ for $n \in \mathbb{N}$.

2. The integrals

$$I_n = \int_0^{\pi/2} \sin^n t \, dt \quad (\text{for } n \in \mathbb{N})$$

can be computed easily via the recurrence formula $nI_n = (n-1)I_{n-2}$ (where $n \ge 2$). Integrate the inequalities $\sin^{2n+1} x \le \sin^{2n} x \le \sin^{2n-1} x$ over $[0, \pi/2]$ to infer *Wallis' formula*,

$$\frac{\pi}{2} = \lim_{n\to\infty} \left[\frac{2 \cdot 2}{1 \cdot 3} \cdot \frac{4 \cdot 4}{3 \cdot 5} \cdots \frac{2n \cdot 2n}{(2n-1)\cdot(2n+1)} \right].$$

Remark. An alternative proof of this formula follows from Corollary 2.2.5, by noticing that $\pi/2 = (\Gamma(1/2))^2/2$.

3. Establish the formula

$$B(x,y) = 2\int_0^{\pi/2} \sin^{2x-1} t \cdot \cos^{2y-1} t\, dt \quad \text{for } x, y > 0,$$

and infer from it that

$$\int_0^{\pi/2} \sin^{2n} t\, dt = \frac{(2n)!\,\pi}{2^{2n+1}(n!)^2} \quad \text{for } n \in \mathbb{N}.$$

4. Use Corollary 2.2.5 to prove Weierstrass' formula,

$$\Gamma(x) = \frac{\mathrm{e}^{-\gamma x}}{x} \prod_{n=1}^{\infty} \Big(1 + \frac{x}{n}\Big)^{-1} \mathrm{e}^{x/n},$$

where $\gamma = \lim_{n\to\infty} (1 + \frac{1}{2} + \cdots + \frac{1}{n} - \log n) = 0.57722\ldots$ is Euler's constant.

5. (The Raabe integral) Prove that

$$\Gamma\Big(\frac{1}{p}\Big)\Gamma\Big(\frac{2}{p}\Big)\cdots\Gamma\Big(\frac{p-1}{p}\Big) = \frac{(2\pi)^{p-1/2}}{p^{1/2}} \quad \text{for all } p \in \mathbb{N}^{\star}.$$

Then infer the integral formula

$$\int_x^{x+1} \log\Gamma(t)\, dt = x(\log x - 1) + \frac{1}{2}\log 2\pi \quad \text{for all } x \geq 0.$$

[*Hint*: Notice that $\int_x^{x+1} \log\Gamma(t)\, dt - x(\log x - 1)$ is constant. The value at $x = 0$ can be computed by using Riemann sums.]

6. (L. Euler) Prove the formula

$$\int_0^{\infty} \frac{t^{x-1}}{1+t}\, dt = \frac{\pi}{\sin \pi x} \quad \text{for } 0 < x < 1.$$

[*Hint*: Put $t = u/(1-u)$ and apply Theorem 2.2.12 (iii).]

7. (An alternative proof of the log-convexity of Γ) Prove the formula

$$\frac{d^2}{dx^2}\log\Gamma(x) = \sum_{n=0}^{\infty} \frac{1}{(x+n)^2} \quad \text{for } x > 0.$$

8. (F. John's approach of the Bohr–Mollerup theorem) Let g be a real-valued concave function on $(0,\infty)$ such that $g(x)/x \to 0$ as $x \to \infty$.

(i) Prove that the difference equation

$$f(x+1) - f(x) = g(x)$$

has one and only one convex solution $f\colon (0,\infty) \to \mathbb{R}$ with $f(1) = 0$, and this solution is given by the formula

$$f(x) = -g(x) + x \cdot \lim_{n\to\infty} \Big(g(n) - \sum_{k=1}^{n-1} \frac{g(x+k) - g(k)}{x} \Big).$$

(ii) (A Stirling type formula) Prove the existence of the limit

$$c = \lim_{x\to\infty} \Big(f(x) + g(x) - \int_{1/2}^{x+1/2} g(t)\, dt \Big).$$

Remark. The Bohr–Mollerup theorem concerns the case where $g = \log$ and $f = \log\Gamma$.

9. (E. Artin [10]) Let U be an open convex subset of $\mathbb{R}^n$ and let μ be a Borel measure on an interval I. Consider the integral transform

$$F(x) = \int_I K(x,t)\, d\mu(t),$$

where the kernel $K(x,t)\colon U \times I \to [0,\infty)$ satisfies the following two conditions:

(i) $K(x,t)$ is μ-integrable in t for each fixed x;
(ii) $K(x,t)$ is log-convex in x for each fixed t.

Prove that F is log-convex on U.
[*Hint*: Apply the Rogers–Hölder inequality, noticing that

$$K((1-\lambda)x + \lambda y, t) \le (K(x,t))^{1-\lambda}(K(y,t))^{\lambda}. \quad]$$

Remark. The *Laplace transform* of a function $f \in L^1(0,\infty)$ is given by the formula $(\mathcal{L}f)(x) = \int_0^\infty f(t)\mathrm{e}^{-tx}\, dt$. By Exercise 9, the Laplace transform of any nonnegative function is log-convex. In the same way one can show that the moment $\mu_\alpha = \int_0^\infty t^\alpha f(t)\, dt$, of any random variable with probability density f, is a log-convex function in α (on each subinterval of $[0,\infty)$ where it is finite).

2.3 Generalities on Multiplicatively Convex Functions

The class of multiplicatively convex functions can be easily described as being constituted by those functions f (acting on subintervals of $(0,\infty)$) such that $\log f(x)$ is a convex function of $\log x$:

Lemma 2.3.1 *Suppose that* $f\colon I \to (0,\infty)$ *is a function defined on a subinterval of* $(0,\infty)$. *Then* f *is multiplicatively convex if and only if*

$$\begin{vmatrix} 1 & \log x_1 & \log f(x_1) \\ 1 & \log x_2 & \log f(x_2) \\ 1 & \log x_3 & \log f(x_3) \end{vmatrix} \geq 0$$

for all $x_1 \leq x_2 \leq x_3$ *in* I; *equivalently, if and only if*

$$f(x_1)^{\log x_3} f(x_2)^{\log x_1} f(x_3)^{\log x_2} \geq f(x_1)^{\log x_2} f(x_2)^{\log x_3} f(x_3)^{\log x_1}$$

for all $x_1 \leq x_2 \leq x_3$ *in* I.

This is nothing but the translation (via Lemma 2.1.1) of the result of Lemma 1.3.2.

In the same spirit, we can show that every multiplicatively convex function $f\colon I \to (0, \infty)$ has finite lateral derivatives at each interior point of I (and the set of all points where f is not differentiable is at most countable). As a consequence, every multiplicatively convex function is continuous in the interior of its domain of definition. Under the presence of continuity, the multiplicative convexity can be restated in terms of geometric mean:

Theorem 2.3.2 *Suppose that* I *is a subinterval of* $(0, \infty)$. *A continuous function* $f\colon I \to (0, \infty)$ *is multiplicatively convex if and only if*

$$x, y \in I \text{ implies } f(\sqrt{xy}) \leq \sqrt{f(x)f(y)}.$$

Proof. The necessity is clear. The sufficiency part follows from the connection between the multiplicative convexity and the usual convexity (as noted in Lemma 2.1.1) and the fact that midpoint convexity is equivalent to convexity in the presence of continuity. See Theorem 1.1.3. □

Theorem 2.3.2 reveals the essence of multiplicative convexity as being the *convexity according to the geometric mean*; in fact, under the presence of continuity, the multiplicatively convex functions are precisely those functions $f\colon I \to (0, \infty)$ for which

$$x_1, \ldots, x_n \in I \text{ implies } f(\sqrt[n]{x_1 \cdots x_n}) \leq \sqrt[n]{f(x_1) \cdots f(x_n)}.$$

In this respect, it is natural to call a function $f\colon I \to (0, \infty)$ *multiplicatively concave* if $1/f$ is multiplicatively convex, and *multiplicatively affine* if f is of the form Cx^α for some $C > 0$ and some $\alpha \in \mathbb{R}$.

A refinement of the notion of multiplicative convexity is that of *strict multiplicative convexity*, which in the context of continuity will mean

$$f(\sqrt[n]{x_1 \cdots x_n}) < \sqrt[n]{f(x_1) \cdots f(x_n)}$$

unless $x_1 = \cdots = x_n$. Clearly, Lemma 2.1.1 (which relates the multiplicatively convex functions and the usual convex functions) has a “strict” counterpart.

A large class of strictly multiplicatively convex functions is indicated by the following result:

Proposition 2.3.3 (G. H. Hardy, J. E. Littlewood and G. Pólya [99, Theorem 177, p. 125]) *Every polynomial $P(x)$ with nonnegative coefficients is a multiplicatively convex function on $(0,\infty)$. More generally, every real analytic function $f(x)=\sum_{n=0}^{\infty} c_n x^n$ with nonnegative coefficients is a multiplicatively convex function on $(0,R)$, where R denotes the radius of convergence.*

Moreover, except for the case of functions Cx^n (with $C>0$ and $n\in\mathbb{N}$), the above examples exhibit strictly multiplicatively convex functions (which are also increasing and strictly convex). In particular,

- exp, sinh and cosh on $(0,\infty)$;
- tan, sec, csc and $\frac{1}{x}-\cot x$ on $(0,\pi/2)$;
- arcsin on $(0,1]$;
- $-\log(1-x)$ and $\frac{1+x}{1-x}$ on $(0,1)$.

See the table of series in I. S. Gradshteyn and I. M. Ryzhik [89].

Proof. By continuity, it suffices to prove only the first assertion. Suppose that $P(x)=\sum_{n=0}^{N} c_n x^n$. According to Theorem 2.3.2, we have to prove that

$$x,y>0 \text{ implies } (P(\sqrt{xy}))^2 \le P(x)P(y),$$

or, equivalently,

$$x,y>0 \text{ implies } (P(xy))^2 \le P(x^2)P(y^2).$$

The later implication is an easy consequence of Cauchy–Buniakovski–Schwarz inequality. $\square$

The following result collects a series of useful remarks for proving the multiplicative convexity of concrete functions:

Lemma 2.3.4

(i) *If a function is* log*-convex and increasing, then it is strictly multiplicatively convex.*
(ii) *If a function f is multiplicatively convex, then the function $1/f$ is multiplicatively concave (and vice versa).*
(iii) *If a function f is multiplicatively convex, increasing and one-to-one, then its inverse is multiplicatively concave (and vice versa).*
(iv) *If a function f is multiplicatively convex, so is $x^{\alpha}[f(x)]^{\beta}$ (for all $\alpha\in\mathbb{R}$ and all $\beta>0$).*
(v) *If f is continuous, and one of the functions $f(x)^x$ and $f(\mathrm{e}^{1/\log x})$ is multiplicatively convex, then so is the other.*

In many cases the inequalities based on multiplicative convexity are better than the direct application of the usual inequalities of convexity (or yield complementary information). This includes the multiplicative analogue of the Hardy–Littlewood–Pólya inequality of majorization:

Proposition 2.3.5 *Suppose that* $x_1 \geq x_2 \geq \cdots \geq x_n$ *and* $y_1 \geq y_2 \geq \cdots \geq y_n$ *are two families of numbers in a subinterval* I *of* $(0, \infty)$ *such that*

$$\begin{aligned} x_1 &\geq y_1 \\ x_1 x_2 &\geq y_1 y_2 \\ &\vdots \\ x_1 x_2 \cdots x_{n-1} &\geq y_1 y_2 \cdots y_{n-1} \\ x_1 x_2 \cdots x_n &= y_1 y_2 \cdots y_n. \end{aligned}$$

Then

$$f(x_1) f(x_2) \cdots f(x_n) \geq f(y_1) f(y_2) \cdots f(y_n)$$

for every multiplicatively convex function $f \colon I \to (0, \infty)$.

A result due to H. Weyl [245] (see also [155]) gives us the basic example of a pair of sequences satisfying the hypothesis of Proposition 2.3.5: *Consider a matrix* $A \in \mathrm{M}_n(\mathbb{C})$ *having the eigenvalues* $\lambda_1, \ldots, \lambda_n$ *and the singular numbers* $s_1, \ldots, s_n$*, and assume that they are rearranged such that* $|\lambda_1| \geq \cdots \geq |\lambda_n|$*, and* $s_1 \geq \cdots \geq s_n$*. Then:*

$$\left| \prod_{k=1}^{m} \lambda_k \right| \leq \prod_{k=1}^{m} s_k \quad \text{for } m = 1, \ldots, n-1 \quad \text{and} \quad \left| \prod_{k=1}^{n} \lambda_k \right| = \prod_{k=1}^{n} s_k.$$

Recall that the *singular numbers* of a matrix A are precisely the eigenvalues of its modulus, $|A| = (A^\star A)^{1/2}$; the spectral mapping theorem assures that $s_k = |\lambda_k|$ when A is Hermitian. The fact that all examples come this way was noted by A. Horn; see [155] for details.

According to the discussion above the following result holds:

Proposition 2.3.6 *Let* $A \in \mathrm{M}_n(\mathbb{C})$ *be any matrix having the eigenvalues* $\lambda_1, \ldots, \lambda_n$ *and the singular numbers* $s_1, \ldots, s_n$*, listed such that* $|\lambda_1| \geq \cdots \geq |\lambda_n|$ *and* $s_1 \geq \cdots \geq s_n$*. Then*

$$\prod_{k=1}^{n} f(s_k) \geq \prod_{k=1}^{n} f(|\lambda_k|)$$

for every multiplicatively convex function f *which is continuous on* $[0, \infty)$.

In general it is not true that $|\lambda_k| \leq s_k$ for all k. A counterexample is given by the matrix

$$\begin{pmatrix} 0 & 1 \\ 4 & 0 \end{pmatrix}$$

whose eigenvalues are $\lambda_1 = 2 > \lambda_2 = -2$ and the singular numbers are $s_1 = 4 > s_2 = 1$.

Exercises

1. (C. H. Kimberling [126]) Suppose that P is a polynomial with nonnegative coefficients. Prove that

$$(P(1))^{n-1}P(x_1\cdots x_n) \geq P(x_1)\cdots P(x_n)$$

provided that all x_k are either in $[0,1]$ or in $[1,\infty)$. This fact complements Proposition 2.3.3.

2. (The multiplicative analogue of Popoviciu's inequality) Suppose there is given a multiplicatively convex function $f\colon I \to (0,\infty)$. Infer from Theorem 2.3.5 that

$$f(x)\,f(y)\,f(z)\,f^3(\sqrt[3]{xyz}) \geq f^2(\sqrt{xy})f^2(\sqrt{yz})f^2(\sqrt{zx})$$

for all $x,y,z \in I$. Moreover, for the strictly multiplicatively convex functions the equality occurs only when $x=y=z$.

3. Recall that the inverse sine function is strictly multiplicatively convex on $(0,1]$ and infer the following two inequalities in a triangle ΔABC:

$$\sin\frac{A}{2}\sin\frac{B}{2}\sin\frac{C}{2} < \Big(\sin\Big(\frac{1}{2}\sqrt[3]{ABC}\Big)\Big)^3 < \frac{1}{8}$$
$$\sin A\sin B\sin C < (\sin\sqrt[3]{ABC})^3 < \frac{3\sqrt{3}}{8}$$

unless $A=B=C$.

4. (P. Montel [171]) Let $I \subset (0,\infty)$ be an interval and suppose that f is a continuous and positive function on I. Prove that f is multiplicatively convex if and only if

$$2f(x) \leq k^{\alpha}f(kx) + k^{-\alpha}f(x/k)$$

for all $\alpha \in \mathbb{R}$, $x \in I$, and $k>0$, such that kx and x/k both belong to I.

5. (The multiplicative mean) According to Lemma 2.1.1, the multiplicative analog of the arithmetic mean is

$$\begin{aligned}M_*(f) &= \exp\Big(\frac{1}{\log b-\log a}\int_{\log a}^{\log b}\log f(\mathrm{e}^t)\,dt\Big)\\ &= \exp\Big(\frac{1}{\log b-\log a}\int_a^b \log f(t)\frac{dt}{t}\Big),\end{aligned}$$

that is, the geometric mean of f with respect to the measure dt/t. Notice that

$$M_*(1)=1$$
$$\inf f \leq f \leq \sup f \Rightarrow \inf f \leq M_*(f) \leq \sup f$$
$$M_*(fg) = M_*(f)M_*(g).$$

(i) Let $f\colon [a,b] \to (0,\infty)$ be a continuous function defined on a subinterval of $(0,\infty)$ and let $\varphi\colon J \to (0,\infty)$ be a multiplicatively convex continuous function defined on an interval J which includes the image of f. Prove that

$$\varphi(M_*(f)) \le M_*(\varphi \circ f),$$

which is the multiplicative analogue of Jensen's inequality.

(ii) Suppose that $0 < a < b$ and let $f\colon [a,b] \to (0,\infty)$ be a multiplicatively convex continuous function. Prove the following analogue of Hermite–Hadamard inequality,

$$f(\sqrt{ab}) \le M_*(f) \le \sqrt{f(a)f(b)};$$

the left-hand side inequality is strict unless f is multiplicatively affine, while the right-hand side inequality is strict unless f is multiplicatively affine on each of the subintervals $[a,\sqrt{ab}]$ and $[\sqrt{ab},b]$. These inequalities can be improved following an idea similar to that of Remark 1.9.3:

$$\begin{aligned} f(a^{1/2}b^{1/2}) &\le (f(a^{3/4}b^{1/4})f(a^{1/4}b^{3/4}))^{1/2} \le M_*(f) \\ &\le (f(a^{1/2}b^{1/2}))^{1/2} f(a)^{1/4} f(b)^{1/4} \\ &\le (f(a)f(b))^{1/2}. \end{aligned}$$

(iii) Notice that $M_*(f) = \exp(\frac{b-a}{\log b - \log a})$ for $f = \exp|_{[a,b]}$ $(0 < a < b)$. Then, infer from (ii) the inequalities:

$$\frac{a^{3/4}b^{1/4} + a^{1/4}b^{3/4}}{2} < \frac{b-a}{\log b - \log a} < \frac{1}{2}\Big(\frac{a+b}{2} + \sqrt{ab}\Big)$$

$$\exp\Big(\frac{b-a}{\log b - \log a}\Big) < \frac{e^b - e^a}{b-a}.$$

6. Let $f\colon I \to (0,\infty)$ be a function which is multiplicatively convex or multiplicatively concave and let $a > 0$.

(i) Prove that

$$\Big(\prod_{k=1}^{n} f(a^{k/n})\Big)^{1/n} > \Big(\prod_{k=1}^{n+1} f(a^{k/(n+1)})\Big)^{1/(n+1)} > M_*(f)$$

for all $n = 1,2,3,\dots$ in each of the following two cases:
- $I = [1,a]$ (with $a > 1$) and f is increasing;
- $I = [a,1]$ (with $0 < a < 1$) and f is decreasing.

(ii) Prove that the above inequalities will be reversed in each of the following two cases:
- $I = [1,a]$ (with $a > 1$) and f is decreasing;
- $I = [a,1]$ (with $0 < a < 1$) and f is increasing.

(iii) Illustrate the assertions (i) and (ii) in the case of the functions $1+\log x$ and $\exp x$, for $x \ge 1$, and $\sin(\pi x/2)$ and $\cos(\pi x/2)$, for $x \in (0,1]$.

2.4 Multiplicative Convexity of Special Functions

We start this section by noticing that the indefinite integral of a multiplicatively convex function has the same nature:

Proposition 2.4.1 (P. Montel [171]) *Let $f\colon [0,a) \to [0,\infty)$ be a continuous function which is multiplicatively convex on $(0,a)$. Then*

$$F(x) = \int_0^x f(t)\,dt$$

is also continuous on $[0,a)$ and multiplicatively convex on $(0,a)$.

Proof. Due to the continuity of F, it suffices to show that

$$(F(\sqrt{xy}))^2 \leq F(x)F(y) \quad \text{for all } x, y \in [0,a),$$

which is a consequence of the corresponding inequality at the level of integral sums,

$$\Big[\frac{\sqrt{xy}}{n}\sum_{k=0}^{n-1} f\Big(k\,\frac{\sqrt{xy}}{n}\Big)\Big]^2 \leq \Big[\frac{x}{n}\sum_{k=0}^{n-1} f\Big(k\,\frac{x}{n}\Big)\Big]\Big[\frac{y}{n}\sum_{k=0}^{n-1} f\Big(k\,\frac{y}{n}\Big)\Big],$$

that is, of the inequality

$$\Big[\sum_{k=0}^{n-1} f\Big(k\,\frac{\sqrt{xy}}{n}\Big)\Big]^2 \leq \Big[\sum_{k=0}^{n-1} f\Big(k\,\frac{x}{n}\Big)\Big]\Big[\sum_{k=0}^{n-1} f\Big(k\,\frac{y}{n}\Big)\Big].$$

To see that the later inequality holds, first notice that

$$\Big[f\Big(k\,\frac{\sqrt{xy}}{n}\Big)\Big]^2 \leq \Big[f\Big(k\,\frac{x}{n}\Big)\Big]\Big[f\Big(k\,\frac{y}{n}\Big)\Big]$$

and then apply the Cauchy–Buniakovski–Schwarz inequality. $\square$

According to Proposition 2.4.1, the *logarithmic integral*,

$$\mathrm{Li}(x) = \int_2^x \frac{dt}{\log t}, \quad x \geq 2,$$

is multiplicatively convex. This function is important in number theory. For example, if $\pi(x)$ counts the number of primes p such that $2 \leq p \leq x$, then an equivalent formulation of the Riemann hypothesis is the existence of a function $C\colon (0,\infty) \to (0,\infty)$ such that

$$|\pi(x) - \mathrm{Li}(x)| \leq C(\varepsilon)x^{1/2+\varepsilon} \quad \text{for all } x \geq 2 \text{ and all } \varepsilon > 0.$$

Since the function tan is continuous on $[0, \pi/2)$ and strictly multiplicatively convex on $(0, \pi/2)$, a repeated application of Proposition 2.4.1 shows that the *Lobacevski's function*

$$\mathrm{L}(x) = -\int_0^x \log \cos t \, dt$$

is strictly multiplicatively convex on $(0, \pi/2)$.

Starting with $t/(\sin t)$, (which is strictly multiplicatively convex on $(0, \pi/2]$) and then switching to $(\sin t)/t$, a similar argument leads us to the fact that the *integral sine* function,

$$\mathrm{Si}(x) = \int_0^x \frac{\sin t}{t} \, dt,$$

is strictly multiplicatively concave on $(0, \pi/2]$.

Another striking fact is the following:

Proposition 2.4.2 Γ *is a strictly multiplicatively convex function on* $[1, \infty)$.

Proof. In fact, $\log \Gamma(1 + x)$ is strictly convex and increasing on $(1, \infty)$. Moreover, an increasing strictly convex function of a strictly convex function is strictly convex. Hence, $F(x) = \log \Gamma(1 + \mathrm{e}^x)$ is strictly convex on $(0, \infty)$ and thus $\Gamma(1 + x) = \exp F(\log x)$ is strictly multiplicatively convex on $[1, \infty)$. As $\Gamma(1 + x) = x\Gamma(x)$, we conclude that Γ itself is strictly multiplicatively convex on $[1, \infty)$. $\square$

According to Proposition 2.4.2,

$$\Gamma^3(\sqrt[3]{xyz}) < \Gamma(x)\Gamma(y)\Gamma(z) \quad \text{for all } x, y, z \geq 1$$

except the case where $x = y = z$.

On the other hand, by the multiplicative version of Popoviciu's inequality (Exercise 2, Section 2.3), we infer that

$$\Gamma(x)\Gamma(y)\Gamma(z)\Gamma^3(\sqrt[3]{xyz}) \geq \Gamma^2(\sqrt{xy})\Gamma^2(\sqrt{yz})\Gamma^2(\sqrt{zx})$$

for all $x, y, z \geq 1$; the equality occurs only for $x = y = z$.

Another application of Proposition 2.4.2 is the fact that the function $\Gamma(2x+1)/\Gamma(x+1)$ is strictly multiplicatively convex on $[1, \infty)$. This can be seen by using the Gauss–Legendre duplication formula given by Theorem 2.2.9.

Exercises

1. (D. Gronau and J. Matkowski [90]) Prove the following converse to Proposition 2.4.2: If $f \colon (0, \infty) \to (0, \infty)$ verifies the functional equation

$$f(x + 1) = xf(x),$$

the normalization condition $f(1) = 1$, and f is multiplicatively convex on an interval (a, ∞), for some $a > 0$, then $f = \Gamma$.

2. Let $f\colon I \to (0,\infty)$ be a differentiable function defined on a subinterval I of $(0,\infty)$. Prove that the following assertions are equivalent:

(i) f is multiplicatively convex;
(ii) the function $xf'(x)/f(x)$ is nondecreasing;
(iii) f verifies the inequality

$$\frac{f(x)}{f(y)} \geq \left(\frac{x}{y}\right)^{yf'(y)/f(y)} \quad \text{for all } x, y \in I.$$

A similar statement works for the multiplicatively concave functions. Illustrate this fact by considering the restriction of $\sin(\cos x)$ to $(0, \pi/2)$.

3. The *psi function* (also known as the *digamma function*) is defined by

$$\mathrm{Psi}(x) = \frac{d}{dx}\log\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}, \quad x > 0$$

and it can be represented as

$$\mathrm{Psi}(x) = -\gamma - \int_0^1 \frac{t^{x-1}-1}{1-t}\,dt,$$

where γ is Euler's constant. See [9], [89].

(i) Prove that the function Psi satisfies the functional equation

$$\psi(x+1) = \psi(x) + \frac{1}{x}.$$

(ii) Infer from Proposition 2.4.2 and the preceding exercise the inequality

$$\frac{\Gamma(x)}{\Gamma(y)} \geq \left(\frac{x}{y}\right)^{y\,\mathrm{Psi}(y)} \quad \text{for all } x, y \geq 1.$$

4. Let $f\colon I \to (0,\infty)$ be a twice differentiable function defined on a subinterval I of $(0,\infty)$. Prove that f is multiplicatively convex if and only if it verifies the differential inequality

$$x[f(x)f''(x) - f'^2(x)] + f(x)f'(x) \geq 0 \quad \text{for all } x > 0.$$

Infer that the integral sine function is multiplicatively concave.

2.5 An Estimate of the AM–GM Inequality

Suppose that I is a subinterval of $(0,\infty)$ and that $f\colon I \to (0,\infty)$ is a twice differentiable function. According to Lemma 2.1.1, the values of the parameter $\alpha \in \mathbb{R}$ for which the function

$$\varphi(x) = f(x) \cdot x^{(-\alpha/2)\log x}$$

is multiplicatively convex on I are precisely those for which the function

$$\Phi(x) = \log \varphi(\mathrm{e}^x) = \log f(\mathrm{e}^x) - \frac{\alpha x^2}{2}$$

is convex on $\log(I)$. Since the convexity of Φ is equivalent to $\Phi'' \geq 0$, we infer that φ is multiplicatively convex if and only if $\alpha \leq \alpha(f)$, where

$$\begin{aligned}\alpha(f) &= \inf_{x\in\log(I)} \frac{d^2}{dx^2}\log f(\mathrm{e}^x)\\ &= \inf_{x\in I} \frac{x^2\big(f(x)f''(x) - (f'(x))^2\big) + xf(x)f'(x)}{f(x)^2}.\end{aligned}$$

By considering also the upper bound

$$\beta(f) = \sup_{x\in\log(I)} \frac{d^2}{dx^2}\log f(\mathrm{e}^x),$$

we arrive at the following result:

Lemma 2.5.1 *Under the above hypotheses, we have*

$$\begin{aligned}\exp\Big(\frac{\alpha(f)}{2n^2}\sum_{1\leq j<k\leq n}(\log x_j - \log x_k)^2\Big) &\leq \Big(\prod_{k=1}^n f(x_k)\Big)^{1/n} \Big/ f\Big(\Big(\prod_{k=1}^n x_k\Big)^{1/n}\Big)\\ &\leq \exp\Big(\frac{\beta(f)}{2n^2}\sum_{1\leq j<k\leq n}(\log x_j - \log x_k)^2\Big)\end{aligned}$$

for all $x_1, \ldots, x_n \in I$.

Particularly, for $f(x) = \mathrm{e}^x$, $x \in [A, B]$ (where $0 < A \leq B$), we have $\alpha(f) = A$ and $\beta(f) = B$, and we are led to the following improvement upon the AM–GM inequality:

Lemma 2.5.2 *Suppose that* $0 < A \leq B$ *and* $n \in \mathbb{N}^\star$. *Then*

$$\begin{aligned}\frac{A}{2n^2}\sum_{1\leq j<k\leq n}(\log x_j - \log x_k)^2 &\leq \frac{1}{n}\sum_{k=1}^n x_k - \Big(\prod_{k=1}^n x_k\Big)^{1/n}\\ &\leq \frac{B}{2n^2}\sum_{1\leq j<k\leq n}(\log x_j - \log x_k)^2\end{aligned}$$

for all $x_1, \ldots, x_n \in [A, B]$.

Since

$$\frac{1}{2n^2}\sum_{1\leq j<k\leq n}(\log x_j - \log x_k)^2$$

represents the *variance* of the random variable whose distribution is

$$\begin{pmatrix} \log x_1 & \log x_2 & \ldots & \log x_n \\ 1/n & 1/n & \ldots & 1/n \end{pmatrix},$$

Lemma 2.5.2 reveals the probabilistic character of the AM–GM inequality. Using the usual device to approximate the integrable functions by step functions, we can derive from Lemma 2.5.2 the following more general result:

Theorem 2.5.3 *Let* (Ω, Σ, P) *be a probability space and let* X *be a random variable on this space, taking values in the interval* $[A, B]$*, where* $0 < A \leq B$*. Then*

$$A \leq \frac{\mathcal{E}(X) - \mathrm{e}^{\mathcal{E}(\log X)}}{\mathrm{var}(\log X)} \leq B.$$

Here $\mathcal{E}(Z) = \int_X Z(\omega)\, dP(\omega)$ represents the *mathematical expectation* of the random variable Z, and $\mathrm{var}(Z) = \mathcal{E}\big((Z - \mathcal{E}(Z))^2\big)$ the *variance* of Z.

Exercises

1. (H. Kober; see [166, p. 81]) Suppose that $x_1, \ldots, x_n$ are distinct positive numbers, and $\lambda_1, \ldots, \lambda_n$ are positive numbers such that $\lambda_1 + \cdots + \lambda_n = 1$. Prove that

$$\frac{A(x_1, \ldots, x_n\,; \lambda_1, \ldots, \lambda_n) - G(x_1, \ldots, x_n\,; \lambda_1, \ldots, \lambda_n)}{\sum_{i<j}(\sqrt{x_i} - \sqrt{x_j})^2}$$

lies between $\inf_i \lambda_i/(n-1)$ and $\sup_i \lambda_i$.

2. (P. H. Diananda; see [166, p. 83]) Under the same hypothesis as in the precedent exercise, prove that

$$\frac{A(x_1, \ldots, x_n\,; \lambda_1, \ldots, \lambda_n) - G(x_1, \ldots, x_n\,; \lambda_1, \ldots, \lambda_n)}{\sum_{i<j}\lambda_i\lambda_j(\sqrt{x_i} - \sqrt{x_j})^2}$$

lies between $1/(1 - \inf_i \lambda_i)$ and $1/\inf_i \lambda_i$.

3. Suppose that $x_1, \ldots, x_n$ and $\lambda_1, \ldots, \lambda_n$ are positive numbers for which $\lambda_1 + \cdots + \lambda_n = 1$. Put $A_n = A(x_1, \ldots, x_n\,; \lambda_1, \ldots, \lambda_n)$ and $G_n = G(x_1, \ldots, x_n\,; \lambda_1, \ldots, \lambda_n)$.

(i) Compute the integral

$$J(x, y) = \int_0^\infty \frac{t\,dt}{(1+t)(x+yt)^2}.$$

(ii) Infer that $A_n/G_n = \exp(\sum_{k=1}^n \lambda_k (x_k - A_n)^2 J(x_k, A_n))$.

2.6 (M, N)-Convex Functions

The four algebraic variants of convexity we considered in the preceding sections can be embedded into a more general framework, by taking two regular means M and N (on the intervals I and J respectively) and calling a function $f\colon I \to J$ to be (M, N)-*midpoint convex* if it satisfies

$$f(M(x, y)) \leq N(f(x), f(y))$$

for all $x, y \in I$. As noticed in the Introduction, if f is continuous, this yields the (M, N)-convexity of f, that is,

$$f(M(x, y\,; 1-\lambda, \lambda)) \leq N(f(x), f(y)\,; 1-\lambda, \lambda)$$

for all $x, y \in I$ and all $\lambda \in [0, 1]$. The sundry notions such as (M, N)-strict convexity and (M, N)-concavity can be introduced in a natural way.

Many important results, such as the left-hand side of the Hermite–Hadamard inequality and the Jensen inequality, extend to this framework. See Theorems A, B and C in the Introduction.

Other results, like Lemma 2.1.1, can be extended only in the context of quasi-arithmetic means:

Lemma 2.6.1 (J. Aczél [2]) *If φ and ψ are two continuous and strictly monotonic functions (on intervals I and J respectively) and ψ is increasing, then a function $f\colon I \to J$ is $(M_{[\varphi]}, M_{[\psi]})$-convex if and only if $\psi \circ f \circ \varphi^{-1}$ is convex on $\varphi(I)$ in the usual sense.*

Proof. In fact, f is $(M_{[\varphi]}, M_{[\psi]})$-convex if and only if

$$\psi\big(f\big(\varphi^{-1}((1-\lambda)u + \lambda v)\big)\big) \leq (1-\lambda)\psi\big(f(\varphi^{-1}(u))\big) + \lambda\psi\big(f(\varphi^{-1}(v))\big)$$

for all $u, v \in \varphi(I)$ and $\lambda \in [0, 1]$. $\square$

A nice illustration of Lemma 2.6.1 was recently given by D. Borwein, J. Borwein, G. Fee and R. Girgensohn [35], who proved that the volume $V_n(p)$ of the ellipsoid $\{x \in \mathbb{R}^n \mid \|x\|_{L^p} \leq 1\}$ is (H, G)-strictly concave as a function of p:

Theorem 2.6.2 *Given $\alpha > 1$, the function $V_\alpha(p) = 2^\alpha \frac{\Gamma(1+1/p)^\alpha}{\Gamma(1+\alpha/p)}$ verifies the inequality*

$$V_\alpha^{1-\lambda}(p) V_\alpha^{\lambda}(q) < V_\alpha\left(\frac{1}{\frac{1-\lambda}{p} + \frac{\lambda}{q}}\right),$$

for all $p, q > 0$, $p \neq q$ and all $\lambda \in (0, 1)$.

Proof. According to Lemma 2.6.1 it suffices to prove that the function

$$U_\alpha(x) = -\log(V_\alpha(1/x)/2^\alpha) = \log\Gamma(1+\alpha x) - \alpha\log\Gamma(1+x)$$

is strictly convex on $(0,\infty)$ for every $\alpha > 1$. Using the psi function,

$$\mathrm{Psi}(x) = \frac{d}{dx}\log\Gamma(x),$$

we have

$$U''_\alpha(x) = \alpha^2\frac{d}{dx}\mathrm{Psi}(1+\alpha x) - \alpha\frac{d}{dx}\mathrm{Psi}(1+x).$$

Then $U''_\alpha(x) > 0$ on $(0,\infty)$ means $(x/\alpha)U''_\alpha(x) > 0$ on $(0,\infty)$, and the latter holds if the function $x \to x\frac{d}{dx}\mathrm{Psi}(1+x)$ is strictly increasing. Or, according to [9], [89],

$$\frac{d}{dx}\mathrm{Psi}(1+x) = \int_0^\infty \frac{u\mathrm{e}^{ux}}{\mathrm{e}^u - 1}\,du,$$

and an easy computation shows that

$$\frac{d}{dx}\Big(x\frac{d}{dx}\mathrm{Psi}(1+x)\Big) = \int_0^\infty \frac{u[(u-1)\mathrm{e}^u + 1]\mathrm{e}^{ux}}{(\mathrm{e}^u-1)^2}\,du > 0.$$

The result now follows. $\square$

As stated in [35, p. 634], the volume function $V_n(p)$ is neither convex nor concave for $n \geq 3$.

In the next chapter we shall encounter the class of M_p-convex functions $(-\infty \leq p \leq \infty)$. A function $f\colon I \to \mathbb{R}$ is said to be M_p-*convex* if

$$f((1-\lambda)x + \lambda y) \leq M_p(f(x), f(y)\,; 1-\lambda, \lambda)$$

for all $x, y \in I$ and all $\lambda \in [0,1]$ (that is, f is (A, M_p)-convex). In order to avoid trivial situations, the theory of M_p-convex functions is usually restricted to nonnegative functions when $p \in \mathbb{R}$, $p \neq 1$.

The case $p = 1$ corresponds to the usual convex functions, while for $p = 0$ we retrieve the log-convex functions. The case $p = \infty$ is that of *quasiconvex* functions, that is, of functions $f\colon I \to \mathbb{R}$ such that

$$f((1-\lambda)x + \lambda y) \leq \sup\{f(x), f(y)\}$$

for all $x, y \in I$ and all $\lambda \in [0,1]$. Clearly, a function $f\colon I \to \mathbb{R}$ is quasiconvex if and only if its sublevel sets $\{x \mid f(x) \leq \alpha\}$ are convex for all $\alpha \in \mathbb{R}$.

If $p > 0$ (or $p < 0$), a function f is M_p-convex if and only if f^p is convex (or concave, respectively). According to Exercise 8, Section 1.1,

$$M_p(x, y\,; 1-\lambda, \lambda) \leq M_q(x, y\,; 1-\lambda, \lambda) \quad \text{for } -\infty \leq p \leq q \leq \infty,$$

which shows that every M_p-convex function is also M_q-convex for all $q \geq p$.

Exercises

1. Suppose that I and J are nondegenerate intervals and $p, q, r \in \mathbb{R}$, $p < q$. Prove that for every function $f : I \to J$ the following two implications hold true:

- If f is (M_q, M_r)-convex and increasing, then it is also (M_p, M_r)-convex;
- If f is (M_p, M_r)-convex and decreasing, then it is also (M_q, M_r)-convex.

Conclude that the function $V_\alpha(p)$ of Theorem 2.6.2 is also (A, G)-concave and (H, A)-concave.

2. Suppose that M and N are two regular means (respectively on the intervals I and J) and the function $N(\cdot, 1)$ is concave. Prove that:

(i) for every two (M, N)-convex functions $f, g : I \to J$, the function $f+g$ is (M, N)-convex;

(ii) for every (M, N)-convex function $f : I \to J$ and $\alpha > 0$, the function αf is (M, N)-convex.

3. Suppose that $f : I \to \mathbb{R}$ is a continuous function which is differentiable on $\operatorname{int} I$. Prove that f is quasiconvex if and only if for each $x, y \in \operatorname{int} I$,

$$f(y) \leq f(x) \text{ implies } f'(x)(y - x) \leq 0.$$

4. (K. Knopp and B. Jessen; see [99, p. 66]) Suppose that φ and ψ are two continuous functions defined in an interval I such that φ is strictly monotonic and ψ is increasing.

(i) Prove that

$$M_{[\varphi]}(x_1, \ldots, x_n; \lambda_1, \ldots, \lambda_n) = M_{[\psi]}(x_1, \ldots, x_n; \lambda_1, \ldots, \lambda_n)$$

for every family $x_1, \ldots, x_n$ of elements of I and every family $\lambda_1, \ldots, \lambda_n$ of nonnegative numbers with $\sum_{k=1}^n \lambda_k = 1$ $(n \in \mathbb{N}^\star)$ if and only if $\psi \circ \varphi^{-1}$ is affine, that is, $\psi = \alpha\varphi + \beta$ for some constants α and β, with $\alpha \neq 0$.

(ii) Infer that any power mean M_p is a mean $M_{[\varphi]}$, where $\varphi(x) = \log x$, if $p = 0$, and $\varphi(x) = (x^p - 1)/p$, if $p \neq 0$.

5. (M. Nagumo, B. de Finetti and B. Jessen; see [99, p. 68]) Let φ be a continuous increasing function on $(0, \infty)$ such that the quasi-arithmetic mean $M_{[\varphi]}$ is positively homogeneous. Prove that $M_{[\varphi]}$ is one of the power means.

[*Hint*: By Exercise 4 (i), we can replace φ by $\varphi - \varphi(1)$, so we may assume that $\varphi(1) = 0$. The same argument yields two functions α and β such that $\varphi(cx) = \alpha(c)\varphi(x) + \beta(c)$ for all $x > 0$, $c > 0$. The condition $\varphi(1) = 0$ shows that $\beta = \varphi$, so for reasons of symmetry,

$$\varphi(cx) = \alpha(c)\varphi(x) + \varphi(c) = \alpha(x)\varphi(c) + \varphi(x).$$

Letting fixed $c \neq 1$, we obtain that α is of the form $\alpha(x) = 1 + k\varphi(x)$ for some constant k. Then φ verifies the functional equation

$$\varphi(xy) = k\varphi(x)\varphi(y) + \varphi(x) + \varphi(y)$$

for all $x > 0$, $y > 0$. When $k = 0$ we find that $\varphi(x) = C \log x$ for some constant C, so $M_{[\varphi]} = M_0$. When $k \neq 0$ we notice that $\chi = k\varphi + 1$ verifies $\chi(xy) = \chi(x)\chi(y)$ for all $x > 0$, $y > 0$. This leads to $\varphi(x) = (x^p - 1)/k$, for some $p \neq 0$, hence $M_{[\varphi]} = M_p$.]

6. (Convexity with respect to Stolarsky's means) One can prove that the exponential function is (L, L)-convex. See Exercise 5 (iii), Section 2.3. Prove that this function is also (I, I)-convex. What can be said about the logarithmic function? Here L and I are respectively the logarithmic mean and the identric mean.

7. (Few affine functions with respect to the logarithmic mean; see [157]) Prove that the only (L, L)-affine functions $f\colon (0, \infty) \to (0, \infty)$ are the constant functions and the linear functions $f(x) = cx$, for $c > 0$. Infer that the logarithmic mean is not a power mean.

2.7 Relative Convexity

The comparison of quasi-arithmetic means is related to convexity via the following result:

Lemma 2.7.1 *Suppose that $\varphi, \psi\colon I \to \mathbb{R}$ are two strictly monotonic continuous functions. If φ is increasing, then*

$$M_{[\psi]} \leq M_{[\varphi]}$$

if and only if $\varphi \circ \psi^{-1}$ is convex.

Lemma 2.7.1 has important consequences. For example, it yields Clarkson's inequalities (which in turn extend the parallelogram law). The following approach (in the spirit of Orlicz spaces) is due to J. Lamperti [136]:

Theorem 2.7.2 *Suppose that $\Phi\colon [0, \infty) \to \mathbb{R}$, is an increasing and continuous function with $\Phi(0) = 0$ and $\Phi(\sqrt{x})$ convex. Consider a σ-finite measure space (X, Σ, μ) and denote by $L^\Phi(X)$ the set of all equivalence classes of all μ-measurable real-valued functions f such that*

$$I_\Phi(f) = \int_X \Phi(|f(x)|)\, d\mu < \infty.$$

If $f + g$ and $f - g$ belong to $L^\Phi(X)$, then

$$I_\Phi(f + g) + I_\Phi(f - g) \geq 2I_\Phi(f) + 2I_\Phi(g). \tag{2.1}$$

If $\Phi(\sqrt{x})$ *is concave and* f *and* g *belong to* $L^\Phi(X)$, *then the reverse inequality is true. If the convexity or concavity of* $\Phi(\sqrt{x})$ *is strict, equality holds in* (2.1) *if and only if* $fg = 0$ *almost everywhere.*

Corollary 2.7.3 (Clarkson's inequalities [57]) *If* $2 \leq p < \infty$, *and* f *and* g *belong to* $L^p(\mu)$, *then*

$$\|f + g\|_{L^p}^p + \|f - g\|_{L^p}^p \geq 2\|f\|_{L^p}^p + 2\|g\|_{L^p}^p.$$

If $0 < p \leq 2$, *then the reverse inequality holds. In either case, if* $p \neq 2$, *equality occurs if and only if* $fg = 0$ *almost everywhere.*

Clarkson's inequalities easily imply the uniform convexity of the spaces $L^p(\mu)$ for $1 < p < \infty$ (see Exercise 2). J. Lamperti applied Corollary 2.7.3 to give the general form of the linear isometries $T\colon L^p(\mu) \to L^p(\mu)$, for $p > 0$, $p \neq 2$.

Clarkson's inequalities are improved on by Hanner's inequalities. See Exercise 7, Section 3.6.

Proof of Theorem 2.7.2. It suffices to prove the following result: *Suppose that* $\Phi\colon [0, \infty) \to \mathbb{R}$ *is a continuous increasing function with* $\Phi(0) = 0$ *and* $\Phi(\sqrt{t})$ *convex. Then*

$$\Phi(|z + w|) + \Phi(|z - w|) \geq 2\Phi(|z|) + 2\Phi(|w|), \tag{2.2}$$

for all $z, w \in \mathbb{C}$, *while if* $\Phi(\sqrt{t})$ *is concave the reverse inequality is true.*

Provided the convexity or concavity is strict, equality holds if and only if $zw = 0$.

In fact, since $\Phi(\sqrt{t})$ is convex, we infer from Lemma 2.7.1 and the parallelogram law the inequality

$$\begin{aligned} \Phi^{-1}\Big\{\frac{\Phi(|z + w|) + \Phi(|z - w|)}{2}\Big\} &\geq \Big\{\frac{|z + w|^2 + |z - w|^2}{2}\Big\}^{1/2} \\ &= (|z|^2 + |w|^2)^{1/2}. \end{aligned} \tag{2.3}$$

On the other hand, the convexity of $\Phi(\sqrt{t})$ and the fact that $\Phi(0) = 0$ yield that $\Phi(\sqrt{t})/t$ is nondecreasing, that is, $t^2/\Phi(t)$ is nonincreasing (respectively decreasing if the convexity is strict). See Theorem 1.3.1. Taking into account the result of Exercise 1, we infer

$$\Phi^{-1}\{\Phi(|z|) + \Phi(|w|)\} \leq (|z|^2 + |w|^2)^{1/2}, \tag{2.4}$$

and thus (2.2) follows from (2.3), (2.4) and the fact that Φ is increasing. When $\Phi(\sqrt{t})$ is strictly convex, we also obtain from Exercise 1 the fact that (2.4) (and thus (2.2)) is strict unless z or w is zero. $\square$

Lemma 2.7.1 leads us naturally to consider the following concept of relative convexity:

Definition 2.7.4 Suppose that f and g are two real-valued functions defined on the same set X, and g is not a constant function. Then f is said to be *convex relative to* g (abbreviated, $g \lhd f$) if

$$\begin{vmatrix} 1 & g(x) & f(x) \\ 1 & g(y) & f(y) \\ 1 & g(z) & f(z) \end{vmatrix} \geq 0,$$

whenever $x, y, z \in X$ with $g(x) \leq g(y) \leq g(z)$.

When X is an interval and g is continuous and increasing, a small computation shows that the condition $g \lhd f$ is equivalent with the convexity of $f \circ g^{-1}$ (on the interval $J = g(I)$).

Examples 2.7.5
Under appropriate assumptions on the domain and the range of the function f, the following statements hold true:

(i) f is convex if and only if $\mathrm{id} \lhd f$;
(ii) f is log-convex if and only if $\mathrm{id} \lhd \log f$;
(iii) f is (G,G)-convex if and only if $\log \lhd \log f$;
(iv) f is (G,A)-convex if and only if $\log \lhd f$.

A more exotic illustration of the concept of relative convexity is the following fact:

$$f \lhd f^{\alpha} \quad \text{for all } f\colon X \to \mathbb{R}_+ \text{ and all } \alpha \geq 1.$$

For example, $\sin \lhd \sin^2$ on $[0, \pi]$, and $|x| \lhd x^2$ on $\mathbb{R}$.

In the context of C^1-differentiable functions, f is convex with respect to an increasing function g if f'/g' is nondecreasing; in the context of C^2-differentiable functions, f is convex with respect to g if and only if $f''/f' \geq g''/g'$ (provided these ratios exist).

It is important to notice that relative convexity is part of comparative convexity. For this we need the integral analogue of quasi-arithmetic mean,

$$M_{[\varphi]}\Big(\mathrm{id}_{[s,t]}; \frac{1}{t-s}\,dx\Big) = \varphi^{-1}\Big(\frac{1}{t-s}\int_s^t \varphi(x)\,dx\Big).$$

In fact, if $g \lhd f$, then

$$\begin{aligned} f\Big(M_{[g]}\Big(\mathrm{id}_{[a,b]}; \frac{1}{b-a}\,dx\Big)\Big) &= f\Big(g^{-1}\Big(\frac{1}{b-a}\int_a^b g(x)\,dx\Big)\Big) \\ &\leq \frac{1}{b-a}\int_a^b f(x)\,dx = M_1(f|_{[a,b]}) \end{aligned}$$

for all $a < b$ in the domain of f and g.

From the above discussion we can infer the following remark due to H. Alzer [7]: Suppose that f is an increasing continuous function (acting on subintervals of $(0, \infty)$) and $1/f^{-1}$ is convex. Then $1/x \lhd f$. As

$M_{[1/x]}(\mathrm{id}_{[a,b]}\,;\frac{1}{b-a}dx)$ coincides with the logarithmic mean $L(a,b)$, it follows that

$$f(L(a,b)) \leq \frac{1}{b-a}\int_a^b f(x)\,dx = M_1(f|_{[a,b]}).$$

We end this section by extending the Hardy–Littlewood–Pólya inequality to the context of relative convexity. Our approach is based on two technical lemmas.

Lemma 2.7.6 *If $f, g\colon X \to \mathbb{R}$ are two functions such that $g \lhd f$, then*

$$g(x) = g(y) \text{ implies } f(x) = f(y).$$

Proof. Since g is not constant, then there must be a $z \in X$ such that $g(x) = g(y) \neq g(z)$. One of the following two cases may occur:

Case 1: $g(x) = g(y) < g(z)$. This yields

$$0 \leq \begin{vmatrix} 1 & g(x) & f(x) \\ 1 & g(x) & f(y) \\ 1 & g(z) & f(z) \end{vmatrix} = (g(z) - g(x))(f(x) - f(y))$$

and thus $f(x) \geq f(y)$. A similar argument gives us the reverse inequality, $f(x) \leq f(y)$.

Case 2: $g(z) < g(x) = g(y)$. This case can be treated in a similar way. □

Lemma 2.7.7 (The generalization of Galvani's Lemma) *If $g \lhd f$ and x, u, v are points of X such that $g(x) \notin \{g(u), g(v)\}$ and $g(u) \leq g(v)$, then*

$$\frac{f(v) - f(x)}{g(v) - g(x)} \geq \frac{f(u) - f(x)}{g(u) - g(x)}.$$

Proof. In fact, the following three cases may occur:

Case 1: $g(x) < g(u) \leq g(v)$. Then

$$\begin{aligned} 0 &\leq \begin{vmatrix} 1 & g(x) & f(x) \\ 1 & g(u) & f(u) \\ 1 & g(v) & f(v) \end{vmatrix} \\ &= (g(u) - g(x))(f(v) - f(x)) - (g(v) - g(x))(f(u) - f(x)) \end{aligned}$$

and the conclusion of Lemma 2.7.7 is clear.

Case 2: $g(u) \leq g(v) < g(x)$. This case can be treated in the same way.

Case 3: $g(u) < g(x) < g(v)$. According to the discussion above we have

$$\begin{aligned} \frac{f(u) - f(x)}{g(u) - g(x)} &= \frac{f(x) - f(u)}{g(x) - g(u)} \leq \frac{f(v) - f(u)}{g(v) - g(u)} \\ &= \frac{f(u) - f(v)}{g(u) - g(v)} \leq \frac{f(x) - f(v)}{g(x) - g(v)} = \frac{f(v) - f(x)}{g(v) - g(x)} \end{aligned}$$

and the proof is now complete. □

Theorem 2.7.8 (The generalization of the Hardy–Littlewood–Pólya inequality) *Let $f, g\colon X \to \mathbb{R}$ be two functions such that $g \lhd f$ and consider points $x_1, \ldots, x_n, y_1, \ldots, y_n \in X$ and weights $p_1, \ldots, p_n \in \mathbb{R}$ such that:*

(i) $g(x_1) \geq \cdots \geq g(x_n)$ *and* $g(y_1) \geq \cdots \geq g(y_n)$;

(ii) $\sum_{k=1}^{r} p_k g(x_k) \leq \sum_{k=1}^{r} p_k g(y_k)$ *for every* $r = 1, \ldots, n$;

(iii) $\sum_{k=1}^{n} p_k g(x_k) = \sum_{k=1}^{n} p_k g(y_k)$.

Then

$$\sum_{k=1}^{n} p_k f(x_k) \leq \sum_{k=1}^{n} p_k f(y_k).$$

Proof. By mathematical induction. The case $n = 1$ is clear. Assuming the conclusion of Theorem 2.7.8 valid for all families of length $n-1$, let us pass to the families of length n. The case where $g(x_k) = g(y_k)$ for some index k can be settled easily by our hypothesis and Lemma 2.7.6. Therefore we may restrict ourselves to the case where $g(x_k) \neq g(y_k)$ for all indices k. By Abel's summation formula,

$$\sum_{k=1}^{n} p_k f(y_k) - \sum_{k=1}^{n} p_k f(x_k) \tag{2.5}$$

equals

$$\begin{aligned}
&\frac{f(y_n)-f(x_n)}{g(y_n)-g(x_n)}\Big(\sum_{i=1}^{n} p_i g(y_i) - \sum_{i=1}^{n} p_i g(x_i)\Big) \\
&\quad + \sum_{k=1}^{n-1}\Big(\frac{f(y_k)-f(x_k)}{g(y_k)-g(x_k)} - \frac{f(y_{k+1})-f(x_{k+1})}{g(y_{k+1})-g(x_{k+1})}\Big)\Big(\sum_{i=1}^{k} p_i g(y_i) - \sum_{i=1}^{k} p_i g(x_i)\Big)
\end{aligned}$$

which, by (iii), reduces to

$$\sum_{k=1}^{n-1}\Big(\frac{f(y_k)-f(x_k)}{g(y_k)-g(x_k)} - \frac{f(y_{k+1})-f(x_{k+1})}{g(y_{k+1})-g(x_{k+1})}\Big)\Big(\sum_{i=1}^{k} p_i g(y_i) - \sum_{i=1}^{k} p_i g(x_i)\Big).$$

According to (ii), the proof will be complete if we show that

$$\frac{f(y_{k+1})-f(x_{k+1})}{g(y_{k+1})-g(x_{k+1})} \leq \frac{f(y_k)-f(x_k)}{g(y_k)-g(x_k)}$$

for all indices k.

In fact, if $g(x_k) = g(x_{k+1})$ or $g(y_k) = g(y_{k+1})$ for some index k, this follows from (i) and Lemmas 2.7.6 and 2.7.7.

When $g(x_k) > g(x_{k+1})$ and $g(y_k) > g(y_{k+1})$ the following two cases may occur:

Case 1: $g(x_k) \neq g(y_{k+1})$. By a twice application of Lemma 2.7.7 we get

$$\begin{aligned}\frac{f(y_{k+1})-f(x_{k+1})}{g(y_{k+1})-g(x_{k+1})} &= \frac{f(x_{k+1})-f(y_{k+1})}{g(x_{k+1})-g(y_{k+1})} \leq \frac{f(x_k)-f(y_{k+1})}{g(x_k)-g(y_{k+1})} \\ &= \frac{f(y_{k+1})-f(x_k)}{g(y_{k+1})-g(x_k)} \leq \frac{f(y_k)-f(x_k)}{g(y_k)-g(x_k)}.\end{aligned}$$

Case 2: $g(x_k) = g(y_{k+1})$. In this case, $g(x_{k+1}) < g(x_k) = g(y_{k+1}) < g(y_k)$, and Lemmas 2.7.6 and 2.7.7 lead us to

$$\begin{aligned}\frac{f(y_{k+1})-f(x_{k+1})}{g(y_{k+1})-g(x_{k+1})} &= \frac{f(x_k)-f(x_{k+1})}{g(x_k)-g(x_{k+1})} \\ &= \frac{f(x_{k+1})-f(x_k)}{g(x_{k+1})-g(x_k)} \leq \frac{f(y_k)-f(x_k)}{g(y_k)-g(x_k)}.\end{aligned}$$

Consequently, (2.5) is a sum of nonnegative terms, and the proof is complete. □

The classical Hardy–Littlewood–Pólya inequality corresponds to the case where X is an interval, g is the identity, and $p_k = 1$ for all k. In this case, the hypothesis (i) can be replaced by the following one:

(i′) $g(x_1) \geq \cdots \geq g(x_n)$,

see Theorem 1.5.4. When X is an interval, g is the identity, and $p_1, \ldots, p_n$ are arbitrary weights, then the result of Theorem 2.7.8 is known as *Fuchs' inequality* [83]. Clearly, Fuchs' inequality implies Corollary 1.4.3 above.

In a similar way, we can extend another important result in majorization theory, the Tomić–Weyl theorem. See Exercise 5.

Exercises

1. (R. Cooper; see [99, p. 84]) Suppose that $\varphi, \psi \colon I \to (0, \infty)$ are two continuous bijective functions. If φ and ψ vary in the same direction and φ/ψ is nonincreasing, then

$$\psi^{-1}\Big(\sum_{k=1}^{n} \psi(x_k)\Big) \leq \varphi^{-1}\Big(\sum_{k=1}^{n} \varphi(x_k)\Big)$$

for every finite family $x_1, \ldots, x_n$ of elements of I.
[*Hint*: If $h(x)/x$ is nonincreasing for $x > 0$, then $h(\sum_{k=1}^{n} x_k) \leq \sum_{k=1}^{n} h(x_k)$ for every finite family $x_1, \ldots, x_n$ of positive numbers. See Section 1.3, Exercise 8.]

2. Infer from Clarkson's inequalities the *uniform convexity* of the spaces $L^p(\mu)$, for $1 < p < \infty$, that is, if x and y are in the unit ball of $L^p(\mu)$, then

$$\inf\Big\{1 - \Big\|\frac{x+y}{2}\Big\| \;\Big|\; \|x-y\| \geq \varepsilon\Big\} > 0 \quad \text{for all } \varepsilon \in (0,2].$$

3. Suppose that $F, g\colon I \to J$ are two continuous functions and g is strictly monotone. Prove that $g \lhd F$ if and only if for every $\alpha \geq 0$ and every $[a, b] \subset I$ the function $F - \alpha g$ attains its maximum either at a or at b.
Remark. This result can be used to prove sharpened versions of the maximum principle for elliptic partial differential operators. See [242].

4. Suppose that $f\colon [0, \pi/2] \to \mathbb{R}$ is a function such that

$$(f(y) - f(z)) \cos x + (f(z) - f(x)) \cos y + (f(x) - f(y)) \cos z \geq 0$$

for all $x \geq y \geq z$ in $[0, \pi/2]$. Prove that

$$f\Big(\frac{\pi}{7}\Big) - f\Big(\frac{2\pi}{7}\Big) + f\Big(\frac{3\pi}{7}\Big) \leq f(0) - f\Big(\frac{\pi}{3}\Big) + f\Big(\frac{\pi}{2}\Big).$$

5. (An extension of the Tomić–Weyl theorem) Suppose that $f, g\colon X \to \mathbb{R}$ are two synchronous functions with $g \lhd f$. Consider points $x_1, \ldots, x_n$, $y_1, \ldots, y_n$ in X and real weights $p_1, \ldots, p_n$ such that:
(i) $g(x_1) \geq \cdots \geq g(x_n)$ and $g(y_1) \geq \cdots \geq g(y_n)$;
(ii) $\sum_{k=1}^{m} p_k g(x_k) \leq \sum_{k=1}^{m} p_k g(y_k)$ for all $m = 1, \ldots, n$.
Prove that

$$\sum_{k=1}^{n} p_k f(x_k) \leq \sum_{k=1}^{n} p_k f(y_k).$$

2.8 Comments

The idea of transforming a nonconvex function into a convex one by a change of variable has a long history. As far as we know, the class of all multiplicatively convex functions was first considered by P. Montel [171] in a beautiful paper discussing the possible analogues of convex functions in n variables. He motivates his study with the following two classical results:

Hadamard's Three Circles Theorem *Let f be an analytical function in the annulus $a < |z| < b$. Then $\log M(r)$ is a convex function of $\log r$, where*

$$M(r) = \sup_{|z|=r} |f(z)|.$$

G. H. Hardy's Mean Value Theorem *Let f be an analytical function in the annulus $a < |z| < b$ and let $p \in [1, \infty)$. Then $\log M_p(r)$ is a convex function of $\log r$, where*

$$M_p(r) = \left(\frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^p \, d\theta\right)^{1/p}.$$

As $\lim_{p\to\infty} M_p(r) = M(r)$, Hardy's aforementioned result implies Hadamard's. It is well known that Hadamard's result is instrumental in deriving the Riesz–Thorin interpolation theorem (see [99]).

The presentation of the class of multiplicatively convex functions (as was done in Sections 2.3 and 2.4) follows C. P. Niculescu [176]. The multiplicative mean (see [178] and Section 2.3, Exercises 5 and 6) provides the right analogue of the arithmetic mean in a fully multiplicative theory of convexity.

The theory of Euler's functions gamma and beta follows the same steps as in E. Artin [10] and R. Webster [243].

As noted by T. Trif [238], the result of Proposition 2.4.2 can be improved: the gamma function is strictly multiplicatively concave on $(0, \alpha]$ and strictly multiplicatively convex on $[\alpha, \infty)$, where $\alpha \approx 0.21609$ is the unique positive solution of the equation $\operatorname{Psi}(x) + x\frac{d}{dx}\operatorname{Psi}(x) = 0$. This fact has a full generalization in the context of (M_p, M_p)-convexity.

The quantum analogue of the gamma function, the *q-gamma function* Γ_q of F. H. Jackson, is defined by

$$\Gamma_q(x) = \frac{(q\,;q)_\infty}{(q^x\,;q)_\infty}\,(1-q)^{1-x} \quad \text{for } x>0 \ \ (0<q<1),$$

where $(a\,;q)_\infty = \prod_{k=0}^\infty (1-aq^k)$. For it, the Bohr–Mollerup theorem has the following form: Γ_q *is the only solution of the functional equation*

$$\Gamma_q(x+1) = \frac{1-q^x}{1-q}\,\Gamma_q(x)$$

which is log-*convex and satisfies* $\Gamma_q(1) = 1$ (see [9]). Γ_q is multiplicatively convex at least on $(2,\infty)$ (see D. Gronau and J. Matkowski [91]).

The well-known inequalities in a triangle ΔABC, such as

$$\sin A + \sin B + \sin C \le 3\sqrt{3}/2 \quad \text{and} \quad \sin A \sin B \sin C \le 3\sqrt{3}/8,$$

can be traced back to an old paper by G. Berkhan [21], from 1907.

R. A. Satnoianu [221] observed that the functions which are convex, multiplicatively convex and increasing are the source of Erdős–Mordell type inequalities in a triangle. Examples of such functions are numerous. See Proposition 2.3.3.

The estimate given in Theorem 2.5.3 for the AM–GM inequality was mentioned in [176].

The general notion of mean was clarified by B. de Finetti [80].

The idea to consider the general notion of (M, N)-convex function (associated to a pair of means) can be traced back to G. Aumann [13]. Important contributions came from J. Aczél [2], [3], J. Matkowski [157], J. Matkowski and J. Rätz [158], [159]. The canonical extension of a mean, as well as Theorems A, B and C in the Introduction, are due to C. P. Niculescu [183].

The result of Exercise 5, Section 2.6, concerning the characterization of the power means among the quasi-arithmetic means, was recently extended by J. Matkowski [157] to the context of strict and homogeneous means which verify some nondegeneracy conditions.

The comparability Lemma 2.7.1 is due to B. Jessen (see [99, p. 75]). The concept of relative convexity can be also traced back to Jessen (see [99, Theorem 92, p. 75]). Later, it was developed by G. T. Cargo [47], N. Elezović and J. Pečarić [71] and many others. The generalization of the classical inequalities of Hardy–Littlewood–Pólya, Fuchs and Tomić–Weyl to the framework of relative convexity follows closely the paper [189] by C. P. Niculescu and F. Popovici.

Recently, M. Bessenyei and Z. Páles [26] have considered a more general concept of relative convexity, which goes back to a result of G. Pólya; see [99, Theorem 123, p. 98]. Given a pair (ω_1, ω_2) of continuous functions on an interval I, such that

$$\begin{vmatrix} \omega_1(x) & \omega_1(y) \\ \omega_2(x) & \omega_2(y) \end{vmatrix} \neq 0 \quad \text{for all } x < y, \tag{2.6}$$

a function $f\colon I \to \mathbb{R}$ is said to be (ω_1, ω_2)-*convex* if

$$\begin{vmatrix} f(x) & f(y) & f(z) \\ \omega_1(x) & \omega_1(y) & \omega_1(z) \\ \omega_2(x) & \omega_2(y) & \omega_2(z) \end{vmatrix} \geq 0$$

for all $x < y < z$ in I. It is proved that the (ω_1, ω_2)-convexity implies the continuity of f at the interior points of I, as well as the integrability on compact subintervals of I.

If I is an open interval, $\omega_1 > 0$ and the determinant in formula (2.6) is positive, then f is (ω_1, ω_2)-convex if and only if the function $f/\omega_1 \circ (\omega_2/\omega_1)^{-1}$ is convex in the usual sense. Under these restrictions, M. Bessenyei and Z. Páles proved a Hermite–Hadamard type inequality. Note that this case of (ω_1, ω_2)-convexity falls under the incidence of relative convexity.

There is much information available nowadays concerning the Clarkson type inequalities, and several applications have been described. Here we just mention that even the general Edmunds–Triebel logarithmic spaces satisfy Clarkson's inequalities: see [191], where some applications and relations to several previous results and references are also presented.

A classical result due to P. Jordan and J. von Neumann asserts that the parallelogram law characterizes Hilbert spaces among Banach spaces. See M. M. Day [64, pp. 151–153]. There are two important generalizations of the parallelogram law (both simple consequences of the inner-product structure).

The Leibniz–Lagrange identity. *Suppose there is given a system of weighted points $(x_1, m_1), \ldots, (x_r, m_r)$ in an inner-product space H, whose barycenter position is*

$$x_G = \sum_{k=1}^{r} m_k x_k \Big/ \sum_{k=1}^{r} m_k.$$

Then for all points $x \in H$ we have the equalities

$$\sum_{k=1}^{r} m_k \|x - x_k\|^2 = \Big(\sum_{k=1}^{r} m_k\Big)\|x - x_G\|^2 + \sum_{k=1}^{r} m_k \|x_G - x_k\|^2$$
$$= \Big(\sum_{k=1}^{r} m_k\Big)\|x - x_G\|^2 + \frac{1}{\sum_{k=1}^{r} m_k} \cdot \sum_{i<j} m_i m_j \|x_i - x_j\|^2.$$

This identity is at the origin of many well-known formulas concerning the distances between some special points in a triangle. For example, in the case where x_1, x_2, x_3 are the vertices of a triangle and m_1, m_2, m_3 are proportional to the length sides a, b, c, then x_G is precisely the center I of the inscribed circle. The above identity gives us (for $x = O$, the center of the circumscribed circle) the celebrated *formula of Euler*,

$$OI^2 = R(R - 2r).$$

More information can be found at `www.neiu.edu/~mathclub/Seminar Notes/ Some Mathematical Consequences of the Law of the Lever`.

E. Hlawka's identity. *We have*

$$\|x\|^2 + \|y\|^2 + \|z\|^2 + \|x + y + z\|^2 = \|x + y\|^2 + \|y + z\|^2 + \|z + x\|^2,$$

for all x, y, z in an inner-product space H.

This yields *Hlawka's inequality*: In any inner-product space H, for all $x, y, z \in H$ we have

$$\|x + y + z\| + \|x\| + \|y\| + \|z\| - \|x + y\| - \|y + z\| - \|z + x\| \geq 0.$$

In fact, based on Hlawka's identity, the left-hand side equals

$$(\|x\| + \|y\| - \|x + y\|)\Big(1 - \frac{\|x\| + \|y\| + \|x + y\|}{\|x\| + \|y\| + \|z\| + \|x + y + z\|}\Big)$$
$$+ (\|y\| + \|z\| - \|y + z\|)\Big(1 - \frac{\|y\| + \|z\| + \|y + z\|}{\|x\| + \|y\| + \|z\| + \|x + y + z\|}\Big)$$
$$+ (\|z\| + \|x\| - \|z + x\|)\Big(1 - \frac{\|z\| + \|x\| + \|z + x\|}{\|x\| + \|y\| + \|z\| + \|x + y + z\|}\Big)$$

which is a combination of nonnegative terms.

Hlawka's inequality is not characteristic to Euclidean spaces! In fact, it was extended by J. Lindenstrauss and A. Pełczyński [146] to all Banach spaces E whose finite dimensional subspaces can be embedded (linearly and isometrically) in suitable spaces $L^p([0,1])$, with $1 \leq p \leq 2$. On the other hand, Hlawka's inequality does not work for all Banach spaces. A counterexample is provided by $\mathbb{C}^2$, endowed with the sup norm, and the vectors $x = (1, -1)$, $y = (i, i)$, $z = (-i, 1)$.

A large generalization of Hlawka's inequality, based on ergodic theory, was given by M. Rădulescu and S. Rădulescu [210].

3

Convex Functions on a Normed Linear Space

Convex functions (and their relatives) provide basic techniques in a series of domains like optimization theory, partial differential equations and geometric inequalities related to isoperimetric problems. They are presented here in the context of real normed linear spaces (most of the time the Euclidean space $\mathbb{R}^n$).

3.1 Convex Sets

The natural domain for a convex function is a convex set. That is why we shall start by recalling some basic facts on convex sets, which should prove useful for understanding the general concept of convexity. *All ambient linear spaces are assumed to be real.*

A subset C of a linear space E is said to be *convex* if it contains the line segment

$$[x, y] = \{(1-\lambda)x + \lambda y \mid \lambda \in [0,1]\}$$

connecting any of its points x and y.

Fig. 3.1 below shows examples of convex and nonconvex sets. Besides these, convex sets in $\mathbb{R}^2$ include line segments, half-spaces, open or closed triangles, or open discs (plus any part of the boundary).

Many other examples can be obtained by considering the following operation with sets:

$$\lambda A + \mu B = \{\lambda x + \mu y \mid x \in A, y \in B\},$$

for $A, B \subset E$ and $\lambda, \mu \in \mathbb{R}$. See Fig. 3.2. One can prove easily that $\lambda A + \mu B$ is convex, provided that A and B are convex and $\lambda, \mu \geq 0$.

A subset A of E is said to be *affine* if it contains the whole line through any two of its points. Algebraically, this means

$$x, y \in A \text{ and } \lambda \in \mathbb{R} \text{ imply } (1-\lambda)x + \lambda y \in A.$$

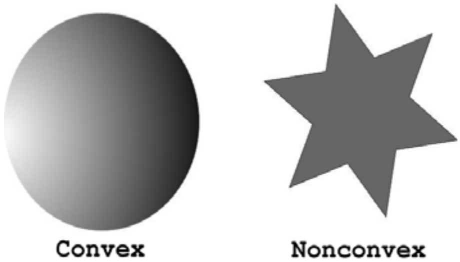

Fig. 3.1. Convex and nonconvex planar sets.

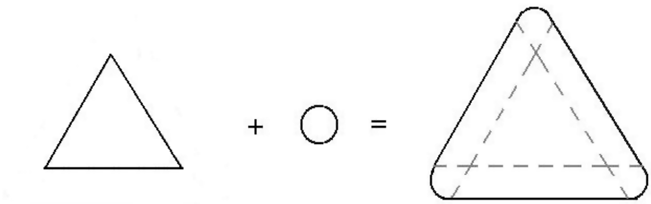

Fig. 3.2. Algebraic sum of sets.

Clearly, any affine subset is also convex (but the converse is not true). It is important to notice that any affine subset A is just the translate of a (unique) linear subspace L (and all translates of a linear space represent affine sets). In fact, for every $a \in A$, the translate

$$L = A - a$$

is a linear space and it is clear that $A = L + a$. For the uniqueness part, notice that if L and M are linear subspaces of E and $a, b \in E$ verify

$$L + a = M + b,$$

then necessarily $L = M$ and $a - b \in L$.

This remark allows us to introduce the concept of *dimension* for an affine set (as the dimension of the linear subspace of which it is a translate).

Given a finite family $x_1, \ldots, x_n$ of points in E, an *affine combination* of them is any point of the form

$$x = \sum_{k=1}^{n} \lambda_k x_k$$

where $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$, and $\sum_{k=1}^{n} \lambda_k = 1$. If, in addition, $\lambda_1, \ldots, \lambda_n \geq 0$, then x is called a *convex combination* (of $x_1, \ldots, x_n$).

Lemma 3.1.1 *A subset C of E is convex (respectively affine) if and only if it contains every convex (respectively affine) combination of points of C.*

Proof. The sufficiency part is clear, while the necessity part can be proved by mathematical induction. See the remark before Lemma 1.1.2. □

Given a subset A of E, the intersection $\mathrm{co}(A)$ of all convex subsets containing A is convex and thus it is the smallest set of this nature containing A. We call it the *convex hull* of A. By using Lemma 3.1.1, one can verify easily that $\mathrm{co}(A)$ consists of all convex combinations of elements of A. The affine variant of this construction yields the *affine hull* of A, denoted $\mathrm{aff}(A)$. As a consequence we can introduce the concept of dimension for convex sets to be the dimension of their affine hulls.

Theorem 3.1.2 (Carathéodory's theorem) *Suppose that A is a subset of a linear space E and its convex hull $\mathrm{co}(A)$ has dimension m. Then each point x of $\mathrm{co}(A)$ is the convex combination of at most $m+1$ points of A.*

Proof. Suppose that $x = \sum_{k=0}^{n} \lambda_k x_k$, where $x_k \in A$, $\lambda_k > 0$ and $\sum_{k=0}^{n} \lambda_k = 1$. If $n > m$, then the set $B = \{x_0, \ldots, x_n\}$ verifies

$$\dim(\mathrm{aff}(B)) \leq \dim(\mathrm{aff}(A)) = m \leq n-1$$

and thus $\{x_1 - x_0, \ldots, x_n - x_0\}$ is a linearly dependent set. This gives us a set of real numbers $\mu_0, \ldots, \mu_n$, not all 0, such that $\sum_{k=0}^{n} \mu_k x_k = 0$ and $\sum_{k=0}^{n} \mu_k = 0$. Choose $t > 0$ for which $\upsilon_k = \lambda_k - t\mu_k \geq 0$ for $k = 0, \ldots, n$ and $\upsilon_j = 0$ for some index j. This allows us to reduce the number of terms in the representation of x. Indeed,

$$x = \sum_{k=0}^{n} \lambda_k x_k = \sum_{k=0}^{n} (\upsilon_k + t\mu_k) x_k = \sum_{k \neq j} \upsilon_k x_k,$$

and $\sum_{k \neq j} \upsilon_k = \sum_{k=0}^{n} \upsilon_k = \sum_{k=0}^{n} (\lambda_k - t\mu_k) = \sum_{k=0}^{n} \lambda_k = 1$. □

The sets of the form $C = \mathrm{co}(\{x_0, \ldots, x_n\})$ are usually called *polytopes*. If $x_1 - x_0, \ldots, x_n - x_0$ are linearly independent, then C is called an *n-simplex* (with vertices $x_0, \ldots, x_n$); in this case, $\dim C = n$. Any point x in an n-simplex C has a unique representation $x = \sum_{k=0}^{n} \lambda_k x_k$, as a convex combination. In this case, the numbers $\lambda_0, \ldots, \lambda_n$ are called the *barycentric coordinates* of x.

An important class of convex sets are the convex cones. A *convex cone* in E is a subset C with the following two properties:

$$C + C \subset C$$
$$\lambda C \subset C \quad \text{for all } \lambda \geq 0.$$

Interesting examples are:

- $\mathbb{R}^n_+ = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_1, \ldots, x_n \geq 0\}$, the nonnegative orthant;
- $\mathbb{R}^n_{++} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_1, \ldots, x_n > 0\}$;
- $\mathbb{R}^n_{\geq} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_1 \geq \cdots \geq x_n\}$;
- $\mathrm{Sym}^+(n, \mathbb{R})$, the set of all positive matrices A of $\mathrm{M}_n(\mathbb{R})$, that is,
$$\langle Ax, x \rangle \geq 0 \quad \text{for all } x \in \mathbb{R}^n;$$
- $\mathrm{Sym}^{++}(n, \mathbb{R})$, the set of all strictly positive matrices A of $\mathrm{M}_n(\mathbb{R})$, that is,
$$\langle Ax, x \rangle > 0 \quad \text{for all } x \in \mathbb{R}^n, \ x \neq 0.$$

They are important because of the orderings they induce:

$$x \leq y \quad \text{if and only if } y - x \in C.$$

So far we have not used any topology; only the linear properties of the space E have played a role.

Suppose now that E is a linear normed space. The following two results relate convexity and topology:

Lemma 3.1.3 *If U is a convex set in a linear normed space, then its interior* $\operatorname{int} U$ *and its closure $\overline{U}$ are convex as well.*

Proof. For example, if $x, y \in \operatorname{int} U$, and $\lambda \in (0, 1)$, then

$$\lambda x + (1 - \lambda)y + u = \lambda(x + u) + (1 - \lambda)(y + u) \in U$$

for all u in a suitable ball $B_\varepsilon(0)$. This shows that $\operatorname{int} U$ is a convex set. Now let $x, y \in \overline{U}$. Then there exist sequences $(x_k)_k$ and $(y_k)_k$ in U, converging to x and y respectively. This yields $\lambda x + (1 - \lambda)y = \lim_{k \to \infty} [\lambda x_k + (1 - \lambda) y_k] \in \overline{U}$ for all $\lambda \in [0, 1]$, that is, $\overline{U}$ is convex as well. $\square$

Notice that affine sets in $\mathbb{R}^n$ are closed because finite dimensional subspaces are always closed.

Lemma 3.1.4 *If U is an open set in a linear normed space E, then its convex hull is open. If E is finite dimensional and K is a compact set, then its convex hull is compact.*

Proof. For the first assertion, let $x = \sum_{k=0}^m \lambda_k x_k$ be a convex combination of elements of the open set U. Then

$$x + u = \sum_{k=0}^m \lambda_k (x_k + u) \quad \text{for all } u \in E$$

and since U is open it follows that $x_k + u \in U$ for all k, provided that $\|u\|$ is small enough. Consequently, $x + u \in \operatorname{co}(U)$ for u in a ball $B_\varepsilon(0)$.

We pass now to the second assertion. Clearly, we may assume that $E = \mathbb{R}^n$. Then consider the map defined by

$$f(\lambda_0, \dots, \lambda_n, x_0, \dots, x_n) = \sum_{k=0}^{n} \lambda_k x_k,$$

where $\lambda_0, \dots, \lambda_n \in [0, 1]$, $\sum_{k=0}^{n} \lambda_k = 1$, and $x_0, \dots, x_n \in K$. Since f is continuous and its domain of definition is a compact space, so is the range of f. According to Carathéodory's theorem, the range of f is precisely $\mathrm{co}(K)$, and this ends the proof. $\square$

While working with a convex subset A of $\mathbb{R}^n$, the natural space containing it is often $\mathrm{aff}(A)$, not $\mathbb{R}^n$, which may be far too large. For example, if $\dim A = k < n$, then A has empty interior. We can talk more meaningfully about the topological notions of interior and boundary by using the notions of relative interior and relative boundary. If A is a convex subset of $\mathbb{R}^n$, the *relative interior* of A, denoted $\mathrm{ri}(A)$, is the interior of A relative to $\mathrm{aff}(A)$. That is, $a \in \mathrm{ri}(A)$ if and only if there is an $\varepsilon > 0$ such that $B_\varepsilon(a) \cap \mathrm{aff}(A) \subset A$. We define the *relative boundary* of A, denoted $\mathrm{rbd}(A)$, as $\mathrm{rbd}(A) = \bar{A} \setminus \mathrm{ri}(A)$. These notions are important in optimization theory; see J. M. Borwein and A. S. Lewis [38].

Exercises

1. Let $S = \{x_0, \dots, x_m\}$ be a finite subset of $\mathbb{R}^n$. Prove that

$$\mathrm{ri}(\mathrm{co}\, S) = \Big\{ \sum_{k=0}^{m} \lambda_k x_k \;\Big|\; \lambda_k \in (0, 1), \sum_{k=0}^{m} \lambda_k = 1 \Big\}.$$

2. Suppose that A and B are convex subsets of $\mathbb{R}^n$, with $A \subset B$ and $\mathrm{aff}(A) = \mathrm{aff}(B)$. Prove that $\mathrm{ri}(a) \subset \mathrm{ri}(B)$.

3. Prove that the relative interior of any nonempty convex subset A of $\mathbb{R}^n$ is dense in A (and thus it is nonempty).

4. (Accessibility lemma) Suppose that A is a convex subset of $\mathbb{R}^n$, $a \in \mathrm{ri}(a)$ and $b \in \bar{A}$. Prove that

$$[a, b) = \{(1 - \lambda)a + \lambda b \mid \lambda \in [0, 1)\} \subset \mathrm{ri}(A).$$

Infer that $\mathrm{ri}(A)$ and $\mathrm{int}(A)$ are also convex.

5. It is well known that all norms on $\mathbb{R}^n$ give rise to the same topology. Prove that all (nonempty) open convex subsets of $\mathbb{R}^n$ are homeomorphic.
[*Hint*: If B is the open unit ball of the Euclidean space $\mathbb{R}^n$, then the mapping $x \to x/(1 - \|x\|^2)$ provides a homeomorphism between B and $\mathbb{R}^n$.]

6. A subset $S = \{x_0, \ldots, x_m\}$ of $\mathbb{R}^n$ is said to be *affinely independent* if the family $\{x_1 - x_0, \ldots, x_m - x_0\}$ is linearly independent. Prove that this means

$$\sum_{k=0}^{m} \lambda_k x_k = 0 \text{ and } \sum_{k=0}^{m} \lambda_k = 0 \text{ imply } \lambda_k = 0 \text{ for all } k \in \{0, \ldots, m\}.$$

Infer that an affinely independent set in $\mathbb{R}^n$ can have at most $n+1$ points.

7. (Helly's theorem; see [213] for applications) Let $(C_i)_{i \in I}$ be a finite collection of convex sets in $\mathbb{R}^n$, where $|I| \geq n+1$. If every subcollection of at most $n+1$ sets has a nonempty intersection, then the entire collection has a nonempty intersection.
[*Hint*: The statement is clear for $|I| = n+1$. Then assume that $|I| > n+1$ and that the statement has already been proved for collections of $|I| - 1$ sets. Choose for each i an element $x_i \in \bigcap_{j \neq i} C_j$. Since $|I| > n+1$ these points are affinely dependent, which yields real scalars λ_i $(i \in I)$, not all null, such that

$$\sum_{i \in I} \lambda_i x_i = 0 \quad \text{and} \quad \sum_{i \in I} \lambda_i = 0.$$

Put $\lambda = \sum_{i \in I} \lambda_i^+ = \sum_{i \in I} \lambda_i^-$. Then $\lambda > 0$ and

$$x = \sum_{i \in I} (\lambda_i^+ / \lambda) x_i = \sum_{i \in I} (\lambda_i^- / \lambda) x_i$$

belongs to C_j for all j. Consider the two cases, $\lambda_j \geq 0$ and $\lambda_j < 0$.]
Remark. Helly's theorem is equivalent with Carathéodory's theorem.

3.2 The Orthogonal Projection

In any normed linear space E we can speak about the *distance* from a point $u \in E$ to a subset $A \subset E$. This is defined by the formula

$$d(u, A) = \inf\{\|u - a\| \mid a \in A\}$$

and represents a numerical indicator of how well u can be approximated by the elements of A. When $E = \mathbb{R}^3$ and A is the x-y plane, the Pythagorean theorem shows that $d(u, A)$ is precisely the distance between u and its orthogonal projection on that plane.

This remark has a notable generalization which will be presented in what follows.

Theorem 3.2.1 *Let C be a nonempty closed convex subset of a Hilbert space H (particularly, of the Euclidean space $\mathbb{R}^n$). Then for each $x \in H$ there is a unique point $P_C(x)$ of C such that*

$$d(x, C) = \|x - P_C(x)\|.$$

We call $P_C(x)$ the *orthogonal projection* of x onto C (or the nearest point of C to x).

Proof. The existence of $P_C(x)$ follows from the definition of the distance from a point to a set and the special geometry of the ambient space. In fact, any sequence $(y_n)_n$ in C such that $\|x - y_n\| \to \alpha = d(x,C)$ is a Cauchy sequence. This is a consequence of the following identity,

$$\|y_m - y_n\|^2 + 4\left\|x - \frac{y_m + y_n}{2}\right\|^2 = 2(\|x - y_m\|^2 + \|x - y_n\|^2)$$

(motivated by the parallelogram law), and the definition of α as an infimum; notice that $\|x - \frac{y_m+y_n}{2}\| \geq \alpha$, which forces $\limsup_{m,n\to\infty} \|y_m - y_n\|^2 = 0$.

Since H is complete, there must exist a point $y \in C$ at which $(y_n)_n$ converges. Then necessarily $d(x,y) = d(x,C)$. The uniqueness of y with this property follows again from the parallelogram law. If y' is another point of C such that $d(x,y') = d(x,C)$ then

$$\|y - y'\|^2 + 4\left\|x - \frac{y + y'}{2}\right\|^2 = 2(\|x - y\|^2 + \|x - y'\|^2)$$

which gives us $\|y - y'\|^2 \leq 0$, a contradiction since it was assumed that the points y and y' are distinct. $\square$

The map $P_C\colon x \to P_C(x)$, from H into itself, is called the *orthogonal projection* associated to C. Clearly,

$$P_C(x) \in C \quad \text{for every } x \in \mathbb{R}^n$$

and

$$P_C(x) = x \quad \text{if and only if } x \in C.$$

In particular,

$$P_C^2 = P_C.$$

P_C is also *monotone*, that is,

$$\langle P_C(x) - P_C(y), x - y\rangle \geq 0 \quad \text{for all } x, y \in H. \tag{3.1}$$

This follows by adding the inequalities $\|x - P_C(x)\|^2 \leq \|x - P_C(y)\|^2$ and $\|y - P_C(y)\|^2 \leq \|y - P_C(x)\|^2$, after replacing the norm by the inner product.

If C is a closed subspace of the Hilbert space H, then P_C is a linear self-adjoint projection and $x - P_C(x)$ is orthogonal on each element of C. This fact is basic for the entire theory of orthogonal decompositions.

Extensions of Theorem 3.2.1 are indicated in Exercise 4 and Theorem C.1.1 (Appendix C).

It is important to reformulate Theorem 3.2.1 in the framework of approximation theory. Suppose that C is a nonempty closed subset in a real linear normed space E. We define the *set of best approximation* from $x \in E$ to C as the set $\mathcal{P}_C(x)$ of all points in C closest to x, that is,

$$\mathcal{P}_C(x) = \{z \in C \mid d(x, C) = \|x - z\|\}.$$

We say that C is a *Chebyshev set* if $\mathcal{P}_C(x)$ is a singleton for all $x \in E$, and a *proximinal set* if all the sets $\mathcal{P}_C(x)$ are nonempty. Theorem 3.2.1 asserts that all nonempty closed convex sets in a Hilbert space are Chebyshev sets. There is an analogue of this theorem valid for the spaces $L^p(\mu)$ $(1 < p < \infty,\ p \neq 2)$, saying that all such sets are proximinal. See Exercise 8 in Section 3.6.

Clearly, the Chebyshev sets are closed. The following result is a partial converse to Theorem 3.2.1:

Theorem 3.2.2 (L. N. H. Bunt) *Every Chebyshev subset of* $\mathbb{R}^n$ *is convex.*

See R. Webster [243, pp. 362–365] for a proof based on Brouwer's fixed point theorem. Proofs based on the differentiability properties of the function $d_C \colon x \to d(x, C)$, are available in the paper by J.-B. Hiriart-Urruty [104], and in the monograph by L. Hörmander [108, pp. 62–63]. They are sketched in Exercise 3, Section 3.8, and Exercise 2, Section 3.11.

V. Klee raised the question whether Theorem 3.2.2 is valid for all real Hilbert spaces. The answer is known to be positive for all Chebyshev sets C such that the map d_C^2 is differentiable. See [104] for details (and an account of Klee's problem).

Exercises

1. Find an explicit formula for the orthogonal projection P_C when C is a closed ball $\overline{B}_r(a)$ in $\mathbb{R}^n$.

2. Let $\ell^\infty(2, \mathbb{R})$ be the space $\mathbb{R}^2$ endowed with the sup norm, $\|(x_1, x_2)\| = \sup\{|x_1|, |x_2|\}$, and let C be the set of all vectors (x_1, x_2) such that $x_2 \geq x_1 \geq 0$. Prove that C is a nonconvex Chebyshev set.

3. Consider in $\mathbb{R}^2$ the nonconvex set $C = \{(x_1, x_2) \mid x_1^2 + x_2^2 \geq 1\}$. Prove that all points of $\mathbb{R}^2$, except the origin, admit a unique closest point in C.

4. (Lions–Stampacchia theorem on a-projections) Let H be a real Hilbert space and let $a \colon H \times H \to \mathbb{R}$ be a coercive continuous bilinear form. Coercivity is meant here as the existence of a positive constant c such that $a(x, x) \geq c\|x\|^2$ for all $x \in H$. Prove that for each $x \in H$ and each nonempty closed convex subset C of H there exists a unique point v in C (called the *a-projection of x onto C*) such that

$$a(x - v, y - v) \leq 0 \quad \text{for all } y \in C.$$

Remark. This theorem has important applications in partial differential equations and optimization theory. See, for example, [14], [69], [70].

3.3 Hyperplanes and Separation Theorems

The notion of a hyperplane represents a natural generalization of the notion of a line in $\mathbb{R}^2$ or a plane in $\mathbb{R}^3$. Hyperplanes are useful to split the whole space into two pieces (called half-spaces).

A *hyperplane* in a real linear space E is any set of constancy of a nonzero linear functional. In other words, a hyperplane is a set of the form

$$H = \{x \in E \mid h(x) = \alpha\}, \tag{3.2}$$

where $h\colon E \to \mathbb{R}$ is a suitable nonzero linear functional and α is a suitable scalar. In this case the sets

$$\{x \in E \mid h(x) \leq \alpha\} \quad \text{and} \quad \{x \in E \mid h(x) \geq \alpha\}$$

are called the *half-spaces* determined by H. We say that H *separates* two sets U and V if they lie in opposite half-spaces (and *strictly separates* U and V if one set is contained in $\{x \in E \mid h(x) < \alpha\}$ and the other in $\{x \in E \mid h(x) \geq \alpha\}$).

When the functional h which appears in the representation formula (3.2) is continuous (that is, when h belongs to the dual space E') we say that the corresponding hyperplane H is *closed*. In the context of $\mathbb{R}^n$, all linear functionals are continuous and thus all hyperplanes are closed. In fact, any linear functional $h\colon \mathbb{R}^n \to \mathbb{R}$ has the form $h(x) = \langle x, z\rangle$, for some $z \in \mathbb{R}^n$ (uniquely determined by h). This follows directly from the linearity of h and the representation of $\mathbb{R}^n$ with respect to the canonical basis:

$$\begin{aligned} h(x) &= h\Big(\sum_{k=1}^{n} x_k e_k\Big) = \sum_{k=1}^{n} x_k h(e_k) \\ &= \langle x, z\rangle, \end{aligned}$$

where $z = \sum_{k=1}^{n} h(e_k)e_k$ is the *gradient* of h.

Some authors define the hyperplanes as the maximal proper affine subsets H of E. Here *proper* means different from E. One can prove that the hyperplanes are precisely the translates of codimension-1 linear subspaces, and this explains the agreement of the two definitions.

The following results on the separation of convex sets by closed hyperplanes are part of a much more general theory that will be presented in Appendix A:

Theorem 3.3.1 (Separation theorem) *Let U and V be two convex sets in a normed linear space E, with $\operatorname{int} U \neq \emptyset$ and $V \cap \operatorname{int} U = \emptyset$. Then there exists a closed hyperplane that separates U and V.*

Theorem 3.3.2 (Strong separation theorem) *Let K and C be two disjoint nonempty convex sets in a normed linear space E with K compact and C closed. Then there exists a closed hyperplane that separates strictly K and C.*

The special case of this result when K is a singleton is known as the *basic separation theorem.*

A proof of Theorems 3.3.1 and 3.3.2 in the finite dimensional case is sketched in Exercises 1 and 2.

Next we introduce the notion of a *supporting hyperplane* to a convex set A in a normed linear space E.

Definition 3.3.3 We say that the hyperplane H *supports* A at a point a in A if $a \in H$ and A is contained in one of the half-spaces determined by H.

Theorem 3.3.1 assures the existence of a supporting hyperplane to any convex set A at a boundary point, provided that A has nonempty interior.

When $E = \mathbb{R}^n$, the existence of a supporting hyperplane of U at a boundary point a will mean the existence of a vector $z \in \mathbb{R}^n$ and of a real number α such that

$$\langle a, z \rangle = \alpha \quad \text{and} \quad \langle x, z \rangle \leq \alpha \quad \text{for all } x \in U.$$

A direct argument for the existence of a supporting hyperplane in the finite dimensional case is given in Exercise 3.

We end this section with a discussion on the geometry of convex sets in finite dimensional spaces.

Definition 3.3.4 Let U be a convex subset of a linear space E. A point z in U is an *extreme point* if it is not an interior point of any linear segment in U, that is, if there do not exist distinct points $x, y \in U$ and numbers $\lambda \in (0, 1)$ such that

$$z = (1 - \lambda)x + \lambda y.$$

The extreme points of a triangle are its vertices. More generally, every polytope $A = \mathrm{co}\{a_0, \ldots, a_m\}$ has finitely many extreme points, and they are among the points $a_0, \ldots, a_m$.

All boundary points of a disc $\overline{D}_R(0) = \{(x, y) \mid x^2 + y^2 \leq R^2\}$ are extreme points; this is an expression of the rotundity of discs. The closed upper half-plane $y \geq 0$ in $\mathbb{R}^2$ has no extreme point.

The extreme points are the landmarks of compact convex sets in $\mathbb{R}^n$:

Theorem 3.3.5 (H. Minkowski) *Every nonempty convex and compact subset K of $\mathbb{R}^n$ is the convex hull of its extreme points.*

Proof. We use induction on the dimension m of K. If $m = 0$ or $m = 1$, that is, when K is a point or a closed segment, the above statement is obvious. Assume the theorem is true for all compact convex sets of dimension at most $m \leq n - 1$. Consider now a compact convex set K whose dimension is $m + 1$ and embed it into a linear subspace E of dimension $m + 1$.

If z is a boundary point of K, then we can choose a supporting hyperplane $H \subset E$ for K through z. The set $K \cap H$ is compact and convex and its dimension is less or equal to m. By the induction hypothesis, z is a convex combination of extreme points of $K \cap H$. Or, any extreme point e of $K \cap H$ is also an extreme point of K. In fact, letting $H = \{t \in E \mid \varphi(t) = \alpha\}$, we may assume that K is included in the half-space $\varphi(t) \leq \alpha$. If $e = (1-\lambda)x + \lambda y$ with $x \neq y$ in K and $\lambda \in (0,1)$, then necessarily $\varphi(x) = \varphi(y) = \alpha$, that is, x and y should be in $K \cap H$, in contradiction with the choice of e.

If z is an interior point of K, then each line through z intersects K in a segment whose endpoints belong to the boundary of K. Consequently, z is a convex combination of boundary points that in turn are convex combinations of extreme points. This ends the proof. $\square$

The result of Theorem 3.3.5 can be made more precise: *every point in a compact convex subset K of $\mathbb{R}^n$ is the convex combination of at most $n+1$ extreme points.* See Theorem 3.1.2.

Exercises

1. Complete the following sketch of the proof of Theorem 3.3.2 in the case when $E = \mathbb{R}^n$: First prove that the distance

$$d = \inf\{\|x - y\| \mid x \in K, y \in C\}$$

is attained for a pair $x_0 \in K$, $y_0 \in C$. Then notice that the hyperplane through x_0, orthogonal to the linear segment $[x_0, y_0]$, determined by x_0 and y_0, has the equation $\langle y_0 - x_0, z - x_0\rangle = 0$. Fix arbitrarily a point $x \in K$. Then $\langle y_0 - x_0, z - x_0\rangle \leq 0$ for every point $z \in [x_0, x]$ (and thus for every $z \in K$). Conclude that every hyperplane through any point inside the segment $[x_0, y_0]$, orthogonal to this segment, separates strictly K and C.

2. Infer the finite dimensional case of Theorem 3.3.1 from Theorem 3.3.2. [*Hint*: It suffices to assume that both sets U and V are closed. Then choose a point $x_0 \in \operatorname{int} U$ and apply the preceding exercise to V and to the compact set

$$K_n = \{x_0 + (1 - 1/n)(x - x_0) \mid x \in U\} \cap \overline{B}_n(0)$$

for $n \in \mathbb{N}^*$. This gives us a sequence of unit vectors u_n and numbers α_n such that $\langle u_n, x\rangle \leq \alpha_n$ for $x \in K_n$ and $\langle u_n, y\rangle \geq \alpha_n$ for $y \in V$. As $(u_n)_n$ and $(\alpha_n)_n$ are bounded, they admit converging subsequences, say to u and α respectively. Conclude that $H = \{z \mid \langle u, z\rangle = \alpha\}$ is the desired separation hyperplane.]

3. (The support theorem) Assume that $E = \mathbb{R}^n$ and a is a point in the relative boundary of the convex subset A of E. Prove that there exists a supporting hyperplane H to A at a which differs from $\operatorname{aff}(A)$.
[*Hint*: We may assume that A is closed, by replacing A with $\overline{A}$. Choose a point $x_0 \in S_1(a) = \{x \mid \|x - a\| = 1\}$ such that

$$d(x_0, A) = \sup\{d(x, A) \mid x \in S_1(a)\},$$

that is, x_0 is the farthest point from A. Notice that a is the point of A closest to x_0 and conclude that the hyperplane $H = \{z \mid \langle x_0 - a, z - a\rangle = 0\}$ supports U at a.

4. Prove that a closed convex set in $\mathbb{R}^n$ is the intersection of closed half-spaces which contain it.

5. A set in $\mathbb{R}^n$ is a *polyhedron* if it is a finite intersection of closed half-spaces. Prove that:

(i) every compact polyhedron is a polytope (the converse is also true);
(ii) every polytope has finitely many extreme points;
(iii) $\mathrm{Sym}^+(2, \mathbb{R})$ is a closed convex cone (with interior $\mathrm{Sym}^{++}(2, \mathbb{R})$) but not a polyhedron.

6. A theorem due to G. Birkhoff (see [243, pp. 246–247]) asserts that every doubly stochastic matrix is a convex combination of permutation matrices. As the set $\Omega_n \subset \mathrm{M}_n(\mathbb{R})$ of all doubly stochastic matrices is compact and convex, and the extreme points of Ω_n are the permutation matrices, Birkhoff's result follows from Theorem 3.3.5.

(i) Verify this fact for $n = 2$.
(ii) Infer from it Rado's characterization of majorization: $x \prec y$ in $\mathbb{R}^n$ if and only if x belongs to the convex hull of the $n!$ permutations of y.

7. Let C be a nonempty subset of $\mathbb{R}^n$. The *polar* set of C, is the set

$$C^\circ = \{x \in \mathbb{R}^n \mid \langle u, x\rangle \le 1 \text{ for every } u \in C\}.$$

(i) Prove that C° is a closed convex set containing 0 and $C \subset D$ implies $D^\circ \subset C^\circ$.
(ii) (The bipolar theorem) Infer from the basic separation theorem that $C^{\circ\circ} = \overline{\mathrm{co}}(C \cup \{0\})$.

3.4 Convex Functions in Higher Dimensions

The notion of a convex function has a natural generalization to real-valued functions defined on an arbitrary convex set.

In what follows U will be a convex set in a real linear space E.

Definition 3.4.1 A function $f\colon U \to \mathbb{R}$ is said to be *convex* if

$$f((1-\lambda)x + \lambda y) \le (1-\lambda)f(x) + \lambda f(y) \tag{3.3}$$

for all $x, y \in U$ and all $\lambda \in [0, 1]$.

The other related notions such as *concave function*, *affine function*, and *strictly convex function* can be introduced as in Section 1.1.

By mathematical induction we can extend the basic inequality (3.3) to the case of arbitrary convex combinations. We shall refer to this as the *discrete case* of Jensen's inequality.

Convexity in the case of several variables is equivalent with convexity on each line segment included in the domain of definition:

Proposition 3.4.2 *A function $f\colon U \to \mathbb{R}$ is convex if and only if for every two points x and y in U the function*

$$\varphi\colon [0,1] \to \mathbb{R}, \quad \varphi(t) = f((1-t)x + ty)$$

is convex.

Notice that convexity of functions in the several variables case means more than convexity in each variable separately; think of the case of the function $f(x,y) = xy$, $(x,y) \in \mathbb{R}^2$, which is not convex, though convex in each variable.

Some simple examples of strictly convex functions on $\mathbb{R}^n$ are as follows:

- $f(x_1,\dots,x_n) = \sum_{k=1}^n \varphi(x_k)$, where φ is a strictly convex function on $\mathbb{R}$.
- $f(x_1,\dots,x_n) = \sum_{i<j} c_{ij}(x_i - x_j)^2$, where the coefficients c_{ij} are positive.
- The distance function $d_U\colon \mathbb{R}^n \to \mathbb{R}$, $d_U(x) = d(x,U)$, associated to a nonempty convex set U in $\mathbb{R}^n$.

We shall next discuss several connections between convex functions and convex sets.

By definition, the *epigraph* of a function $f\colon U \to \mathbb{R}$ is the set

$$\operatorname{epi}(f) = \{(x,y) \mid x \in U, y \in \mathbb{R} \text{ and } f(x) \le y\}.$$

It is easy to verify that $f\colon U \to \mathbb{R}$ is convex if and only if $\operatorname{epi}(f)$ is convex in $E \times \mathbb{R}$. This shows that the *theory of convex functions can be subordinated to the theory of convex sets.*

A practical implication is the existence of supporting hyperplanes for convex functions. To make this more precise, we shall pass to the topological context, where U is an open convex set in a linear normed space E and $f\colon U \to \mathbb{R}$ is a continuous convex function. In this case, $\operatorname{epi}(f)$ has a *nonempty* interior in $E\times\mathbb{R}$ and every point $(a, f(a))$ is a boundary point for $\operatorname{epi}(f)$. According to Theorem 3.3.1, there is a closed hyperplane H in $E\times\mathbb{R}$ that contains $(a, f(a))$ and $\operatorname{epi}(f)$ is contained in one of the half-spaces determined by H. We call this a *supporting hyperplane* to f at a.

The closed hyperplanes H are associated to nonzero continuous linear functionals on $E \times \mathbb{R}$ and the dual space of $E \times \mathbb{R}$ is constituted of all pairs (h,λ), where h is any continuous linear functional on E and λ is any real number. Consequently a supporting hyperplane to f at a is determined by a pair (h,λ) and a real number α such that

$$h(a) + \lambda f(a) = \alpha$$

and

$$h(x) + \lambda y \geq \alpha \text{ for all } y \geq f(x) \quad \text{and all } x \in U.$$

Notice that $\lambda \neq 0$, since otherwise $h(x) \geq h(a)$ for x in a ball $B_r(a)$, which forces $h = 0$. A moment's reflection shows that actually $\lambda > 0$ and thus we are led to the existence of a continuous linear functional h such that

$$f(x) \geq f(a) + h(x - a) \quad \text{for every } x \in U.$$

We call h a *support* of f at a.

From this point we can continue as in the case of functions of one variable, by developing the concept of the subdifferential. We shall come back to this matter in Section 3.7.

We pass now to another connection between convex functions and convex sets.

Given a function $f\colon U \to \mathbb{R}$ and a scalar α, the *sublevel set* L_α of f at height α is the set

$$L_\alpha = \{x \in U \mid f(x) \leq \alpha\}.$$

Lemma 3.4.3 *Each sublevel set of a convex function is a convex set.*

The property of Lemma 3.4.3 characterizes the quasiconvex functions. See Exercise 8.

Convex functions exhibit a series of nice properties related to maxima and minima, which make them important in theoretical and applied mathematics.

Theorem 3.4.4 *Assume that U is a convex subset of a normed linear space E. Then any local minimum of a convex function $f\colon U \to \mathbb{R}$ is also a global minimum. Moreover, the set of global minimizers of f is convex.*

If f is strictly convex in a neighborhood of a minimum point, then the minimum point is unique.

Proof. If a is a local minimum, then for each $x \in U$ there is an $\varepsilon > 0$ such that

$$\begin{aligned} f(a) &\leq f(a + \varepsilon(x - a)) = f((1 - \varepsilon)a + \varepsilon x) \\ &\leq (1 - \varepsilon)f(a) + \varepsilon f(x). \end{aligned} \tag{3.4}$$

This yields $f(a) \leq f(x)$, so a is a global minimum. If f is strictly convex in a neighborhood of a, then the last inequality in (3.4) is strict and the conclusion becomes $f(x) > f(a)$ for all $x \in U$, $x \neq a$. The second assertion is a consequence of Lemma 3.4.3. $\square$

The following result gives us a useful condition for the existence of a global minimum:

Theorem 3.4.5 (K. Weierstrass) *Assume that U is an unbounded closed convex set in $\mathbb{R}^n$ and $f\colon U \to \mathbb{R}$ is a continuous convex function whose sublevel sets are bounded. Then f has a global minimum.*

Proof. Notice that all sublevel sets L_α of f are bounded and closed (and thus compact in $\mathbb{R}^n$). Then every sequence of elements in a sublevel set has a converging subsequence and this yields immediately the existence of global minimizers. □

Under the assumptions of Theorem 3.4.5, the condition on boundedness of sublevel sets is equivalent with the following growth condition:

$$\liminf_{\|x\|\to\infty} \frac{f(x)}{\|x\|} > 0. \tag{3.5}$$

The sufficiency part is clear. For the necessity part, reason by *reductio ad absurdum* and choose a sequence $(x_k)_k$ in U such that $\|x_k\| \to \infty$ and $f(x_k) \leq \|x_k\|/k$. Since the level sets are supposed to be bounded we have $\|x_k\|/k \to \infty$, and this leads to a contradiction. Indeed, for every $x \in U$ the sequence

$$x + \frac{k}{\|x_k\|}\,(x_k - x)$$

is unbounded though lies in some sublevel set $L_{f(x)+\varepsilon}$, with $\varepsilon > 0$.

The functions which verify the condition (3.5) are said to be *coercive*. Clearly, coercivity implies

$$\lim_{\|x\|\to\infty} f(x) = \infty.$$

Convex functions attain their maxima at the boundary:

Theorem 3.4.6 (The maximum principle) *If f is a convex function on a convex subset U of a normed linear space E and attains a global maximum at an interior point of U, then f is constant.*

Proof. Assume that f is not constant and attains a global maximum at the point $a \in \operatorname{int} U$. Choose $x \in U$ such that $f(x) < f(a)$ and $\varepsilon \in (0,1)$ such that $y = a + \varepsilon(a - x) \in U$. Then $a = y/(1+\varepsilon) + \varepsilon x/(1+\varepsilon)$, which yields a contradiction since

$$f(a) \leq \frac{1}{1+\varepsilon} f(y) + \frac{\varepsilon}{1+\varepsilon} f(x) < \frac{1}{1+\varepsilon} f(a) + \frac{\varepsilon}{1+\varepsilon} f(a) = f(a).$$

□

A generalization of the maximum principle is given in Corollary A.3.3 (Appendix A). We end this section with an important consequence of Theorem 3.3.5.

Theorem 3.4.7 *If f is a continuous convex function on a compact convex subset K of $\mathbb{R}^n$, then f attains a global maximum at an extreme point.*

Proof. Assume that f attains its global maximum at $a \in K$. By Theorem 3.3.5, the point a can be represented as a convex combination of extreme points, say $a = \sum_{k=1}^{m} \lambda_k e_k$. Then $f(a) \leq \sum_{k=1}^{m} \lambda_k f(e_k) \leq \sup_k f(e_k)$, which forces $f(a) = f(e_k)$ for some k. $\square$

For functions defined on n-dimensional intervals $[a_1, b_1] \times \cdots \times [a_n, b_n]$ in $\mathbb{R}^n$, Theorem 3.4.7 extends to the case of continuous functions which are convex in each variable (when the others are kept fixed). This fact can be proved by one-variable means (taking into account Theorem 1.1.3). A sample is offered by Exercise 3.

In the infinite dimensional setting, it is difficult to state fairly general results on maximum-attaining. Besides, the deep results of Banach space theory appears to be crucial in answering questions which at first glance may look simple. Here is an example. By the Eberlein–Šmulyan theorem (see Theorem A.1.8, Appendix A) it follows that each continuous linear functional on a reflexive Banach space E achieves its norm on the unit ball. Surprisingly, these are the *only* Banach spaces for which the norm-attaining phenomenon occurs. This was proved by R. C. James (see [64, p. 63]).

Exercises

1. Prove that the general form of an affine function $f\colon \mathbb{R}^n \to \mathbb{R}$ is $f(x) = \langle x, u \rangle + a$, where $u \in \mathbb{R}^n$ and $a \in \mathbb{R}$.

2. (A. Engel [72, p. 177]) A finite set $\mathcal{P}$ of n points ($n \geq 2$) is given in the plane. For any line L, denote by $d(L)$ the sum of distances from the points of $\mathcal{P}$ to the line L. Consider the set $\mathcal{L}$ of the lines L such that $d(L)$ has the lowest possible value. Prove that there exists a line of $\mathcal{L}$ passing through two points of $\mathcal{P}$.

3. Find the maximum of the function

$$f(a, b, c) = \left[3\left(a^5 + b^7 \sin\frac{\pi a}{2} + c\right) - 2(bc + ca + ab)\right]$$

for $a, b, c \in [0, 1]$.
[*Hint*: Notice that $f(a, b, c) \leq \sup[3(a + b + c) - 2(bc + ca + ab)] = 4.$]

4. (i) Prove that the set $\mathrm{Sym}^{++}(n, \mathbb{R})$, of all matrices $A \in \mathrm{M}_n(\mathbb{R})$ which are strictly positive, is open and convex.
(ii) Prove that the function

$$f\colon \mathrm{Sym}^{++}(n, \mathbb{R}) \to \mathbb{R}, f(A) = \log(\det A)$$

is concave.
[*Hint:* (ii) First, notice that $\int_{\mathbb{R}^n} \mathrm{e}^{-\langle Ax, x \rangle}\, dx = \pi^{n/2}/\sqrt{\det A}$ for every A in $\mathrm{Sym}^{++}(n, \mathbb{R})$; there is no loss of generality assuming that A is diagonal. Then, for every $A, B \in \mathrm{Sym}^{++}(n, \mathbb{R})$ and every $\alpha \in (0, 1)$, we have

$$\int_{\mathbb{R}^n} \mathrm{e}^{-\langle [\alpha A+(1-\alpha)B]x,x\rangle}\,dx \le \left(\int_{\mathbb{R}^n} \mathrm{e}^{-\langle Ax,x\rangle}\,dx\right)^{\alpha}\left(\int_{\mathbb{R}^n} \mathrm{e}^{-\langle Bx,x\rangle}\,dx\right)^{1-\alpha},$$

by the Rogers–Hölder inequality. This yields the log-concavity of the function det:

$$\det(\alpha A+(1-\alpha)B) \ge (\det A)^{\alpha}(\det B)^{1-\alpha}. \quad]$$

Extend this formula in the setting of positive matrices (e.g., using perturbations of the form $A+\varepsilon I$ and $B+\varepsilon I$).

5. (The John–Loewner ellipsoid) The aim of this exercise is to sketch a proof of the following result: *Given a compact set A in $\mathbb{R}^n$ with nonempty interior, there exists one and only one ellipsoid E of minimal volume, containing A*; the ellipsoids are supposed to be centered at the origin. Recall that an ellipsoid is a set of the form $C(A)=\{x\in\mathbb{R}^n \mid \langle Ax,x\rangle \le 1\}$, associated to a matrix $A\in \mathrm{Sym}^{++}(n,\mathbb{R})$.

(i) Notice that given two ellipsoids which contain A, there is a third smaller ellipsoid which contains A. Infer from this remark the uniqueness of E.

(ii) Use a compactness argument to derive the existence of an ellipsoid of minimal volume containing A.

(iii) Verify that the volume of $C(A)$ is

$$\mathrm{Vol}_n(C(A)) = \frac{\pi^{n/2}}{\Gamma(n/2+1)}\,(\det A)^{-1/2}.$$

(iv) Infer from the preceding exercise that the function $A\to \mathrm{Vol}_n\,(C(A))$ is strictly convex (and thus it admits at most one minimum).

Remark. The above result has a series of strong consequences. See the paper by M. Berger [23]. We recall here a renorming result due to F. John: *For every n-dimensional real Banach space there is a linear isomorphism $T\colon E\to\mathbb{R}^n$ such that $\|T\|\cdot\|T^{-1}\|\le\sqrt{n}$.*

6. Suppose that $\varphi_1,\ldots,\varphi_n$ are convex functions defined on the same convex set D in $\mathbb{R}^n$ and $f\colon\mathbb{R}^n\to\mathbb{R}$ is a nondecreasing convex function. Prove that $F(x)=f(\varphi_1(x),\ldots,\varphi_n(x))$ is convex on D. Here "nondecreasing" means nondecreasing in each variable (when the others are kept fixed).

7. (i) Prove that the limit of any pointwise converging sequence of convex functions is a convex function.

(ii) Let $(f_\alpha)_\alpha$ be a family of convex functions defined on the same convex set U, such that $f(x)=\sup_\alpha f_\alpha(x)<\infty$ for all $x\in U$. Prove that f is convex.

8. A function $f\colon U\to\mathbb{R}$ defined on a convex set U in $\mathbb{R}^n$ is said to be *quasiconvex* if

$$f((1-\lambda)x+\lambda y)\le \sup\{f(x),f(y)\}$$

for all $x,y\in U$ and all $\lambda\in[0,1]$.

(i) Prove that f is quasiconvex if and only if its sublevel sets L_α are convex for every real number α.

(ii) Extend Theorem 3.4.7 to the context of quasiconvex functions.

9. *Brouwer's fixed point theorem* asserts that any continuous self map of a nonempty compact convex subset of $\mathbb{R}^n$ has a fixed point. See [38, pp. 179–182] for details. The aim of this exercise is to outline a string of results which relates this theorem with the topics of convexity.

(i) Infer from Brouwer's fixed point theorem the following result due to Knaster–Kuratowski–Mazurkiewicz (also known as the *KKM theorem*): Suppose that X is a nonempty subset of $\mathbb{R}^n$ and M is a function which associates to each $x \in X$ a closed nonempty subset $M(x)$ of X. If

$$\mathrm{co}(F) \subset \bigcup_{x \in F} M(x)$$

for every finite subset $F \subset X$, then $\bigcap_{x \in F} M(x) \neq \emptyset$ for every finite subset $F \subset X$. Moreover, $\bigcap_{x \in X} M(x) \neq \emptyset$ if X is compact.
[*Hint*: If $\bigcap_{x \in F} M(x)$ is empty for some finite subset F, then the map

$$y \in \mathrm{co}(F) \to \Big[\sum_{x \in F} d_{M(x)}(y) x\Big] \Big/ \Big[\sum_{x \in F} d_{M(x)}(y)\Big]$$

admits a fixed point z. Letting $G = \{x \in F \mid z \notin M(x)\}$, then z should be in $\mathrm{co}(G)$, and this leads to a contradiction.]

(ii) Prove that the KKM theorem yields the *Ky Fan minimax inequality*: Suppose that C is a nonempty, compact, and convex subset of $\mathbb{R}^n$. If $f\colon C \times C \to \mathbb{R}$, $f = f(x,y)$, is quasiconcave in the first variable and lower semicontinuous in the second variable, then

$$\inf_y \sup_x f(x,y) \leq \sup_x f(x,x).$$

(iii) (Nash equilibrium) Consider the set $C = C_1 \times \cdots \times C_m$, where each set C_k is a nonempty, compact and convex subset of $\mathbb{R}^n$. Consider also continuous functions $f_1, \ldots, f_m\colon C \to \mathbb{R}$ such that, for each k, the function

$$x_k \in C_k \to f(y_1, \ldots, x_k, \ldots, y_m)$$

is convex on C_k for all $y_i \in C_i$, $i \neq k$. Then there exists an element $c = (c_1, \ldots, c_m) \in C$ such that

$$f_k(c) \leq f_k(c_1, \ldots, x_k, \ldots, c_m) \quad \text{for all } x_k \in C_k,\ k \in \{1, \ldots, m\}.$$

[*Hint*: Apply the Ky Fan minimax inequality to the function

$$f(x,y) = \sum_{k=1}^{m} [f_k(y) - f_k(y_1, \ldots, x_k, \ldots, y_m)]. \quad]$$

10. (Multiplicatively convex functions of several variables) Let

$$f(x_1, \dots, x_n) = \sum_{k=1}^{m} a_k x_1^{r_{1k}} \cdots x_n^{r_{nk}}, \quad (x_1, \dots, x_n) \in \mathbb{R}^n_{++}$$

where $a_k > 0$ and $r_{ij} \in \mathbb{R}$. Prove that $g(y_1, \dots, y_n) = \log f(\mathrm{e}^{y_1}, \dots, \mathrm{e}^{y_n})$ is convex on $\mathbb{R}^n$.

3.5 Continuity of Convex Functions

In Section 1.3 we proved that a convex function defined on an open interval is continuous. Here we establish the corresponding results for real-valued functions defined on an open convex set in $\mathbb{R}^n$. The basic remark refers to a local property of convex functions.

Lemma 3.5.1 *Every convex function f defined on an open convex set U in $\mathbb{R}^n$ is locally bounded (that is, each $a \in U$ has a neighborhood on which f is bounded).*

Proof. For $a \in U$ arbitrarily fixed, choose a cube K in U, centered at a, with vertices $v_1, \dots, v_{2^n}$. Clearly, K is a neighborhood of a. Every $x \in K$ is a convex combination of vertices and thus

$$f(x) = f\Big(\sum_{k=1}^{2^m} \lambda_k v_k\Big) \leq M = \sup_{1 \leq k \leq 2^m} f(v_k),$$

so f is bounded above on K. By the symmetry of K, for every $x \in K$ there is a $y \in K$ such that $a = (x+y)/2$. Then $f(a) \leq (f(x) + f(y))/2$, which yields $f(x) \geq 2f(a) - f(y) \geq 2f(a) - M$, and the proof is complete. $\square$

Proposition 3.5.2 *Let f be a convex function on an open convex set U in $\mathbb{R}^n$. Then f is locally Lipschitz. In particular, f is continuous on U.*

According to a theorem due to Rademacher (see Theorem 3.11.1), we can infer from Proposition 3.5.2 that every convex function on an open convex set U in $\mathbb{R}^n$ is almost everywhere differentiable. A direct proof will be given in Section 3.8 (see Theorem 3.8.3).

Proof. According to the preceding lemma, given $a \in U$, we may find a ball $B_{2r}(a) \subset U$ on which f is bounded above, say by M. For $x \neq y$ in $B_r(a)$, put $z = y + (r/\alpha)(y - x)$, where $\alpha = \|y - x\|$. Clearly, $z \in B_{2r}(a)$. As

$$y = \frac{r}{r+\alpha}\, x + \frac{\alpha}{r+\alpha}\, z,$$

from the convexity of f we infer that

$$f(y) \le \frac{r}{r+\alpha} f(x) + \frac{\alpha}{r+\alpha} f(z).$$

Then

$$\begin{aligned} f(y) - f(x) &\le \frac{\alpha}{r+\alpha} [f(z) - f(x)] \\ &\le \frac{\alpha}{r} [f(z) - f(x)] \le \frac{2M}{r} \|y - x\| \end{aligned}$$

and the proof ends by interchanging the roles of x and y. $\square$

Corollary 3.5.3 *Let f be a convex function defined on a convex set A in $\mathbb{R}^n$. Then f is Lipschitz on each compact convex subset of* $\mathrm{ri}(A)$ (*and thus f is continuous on* $\mathrm{ri}(A)$).

Proof. Clearly, we may assume that $\mathrm{aff}(A) = \mathbb{R}^n$. In this case, $\mathrm{ri}(A) = \mathrm{int}(A)$ and Proposition 3.5.2 applies. $\square$

The infinite dimensional analogue of Proposition 3.5.2 is as follows:

Proposition 3.5.4 *Let f be a convex function on an open convex set U in a normed linear space. If f is bounded above in a neighborhood of one point of U, then f is locally Lipschitz on U. In particular, f is a continuous function.*

The proof is similar with that of Proposition 3.5.2, with the difference that the role of Lemma 3.5.1 is taken by the following lemma:

Lemma 3.5.5 *Let f be a convex function on an open convex set U in a normed linear space. If f is bounded above in a neighborhood of one point of U, then f is locally bounded on U.*

Proof. Suppose that f is bounded above by M on a ball $B_r(a)$. Let $x \in U$ and choose $\rho > 1$ such that $z = a + \rho(x - a) \in U$. If $\lambda = 1/\rho$, then

$$V = \{v \mid v = (1-\lambda)y + \lambda z,\ y \in B_r(a)\}$$

is a neighborhood of $x = (1-\lambda)a + \lambda z$, with radius $(1-\lambda)r$. Moreover, for $v \in V$ we have

$$f(v) \le (1-\lambda)f(y) + \lambda f(z) \le (1-\lambda)M + \lambda f(z).$$

To show that f is bounded below in the same neighborhood, choose arbitrarily $v \in V$ and notice that $2x - v \in V$. Consequently, $f(x) \le f(v)/2 + f(2x-v)/2$, which yields $f(v) \ge 2f(x) - f(2x - v) \ge 2f(x) - M$. $\square$

A convex function on an infinite dimensional Banach space E is not necessarily continuous. Actually, one can prove that the only Banach spaces E such that every convex function $f\colon E \to \mathbb{R}$ is continuous are the finite dimensional ones. This is a consequence of the well-known fact that the norm and the weak topology agree only in the finite dimensional case. See [64, Lemma 1, p. 45].

In applications it is often useful to consider extended real-valued functions, defined on a real linear space E.

Definition 3.5.6 A function $f\colon E \to \overline{\mathbb{R}}$ is said to be *convex* if its epigraph,

$$\operatorname{epi}(f) = \{(x, y) \mid x \in E,\ y \in \mathbb{R} \text{ and } f(x) \le y\}$$

is a convex subset of $E \times \mathbb{R}$.

The *effective domain* of a convex function $f\colon E \to \overline{\mathbb{R}}$ is the set

$$\operatorname{dom} f = \{x \mid f(x) < \infty\}.$$

Clearly, this is a convex set. Most of the time we shall deal with *proper convex functions*, that is, with convex functions $f\colon E \to \mathbb{R} \cup \{\infty\}$ which are not identically ∞. In their case, the property of convexity can be reformulated in more familiar terms,

$$f((1-\lambda)x + \lambda y) \le (1-\lambda)f(x) + \lambda f(y)$$

for all $x, y \in E$ and all $\lambda \in [0, 1]$ for which the right hand side is finite.

If U is a convex subset of E, then every convex function $f\colon U \to \mathbb{R}$ extends to a proper convex function $\widetilde{f}$ on E, letting $\widetilde{f}(x) = \infty$ for $x \in E \backslash U$. Another basic example is related to the indicator function. The *indicator function* of a nonempty subset A is defined by the formula

$$\delta_A(x) = \begin{cases} 0 & \text{if } x \in A, \\ \infty & \text{if } x \in E \backslash A. \end{cases}$$

Clearly, A is convex if and only if δ_A is a proper convex function.

The sublevel sets of a proper convex function $f\colon E \to \mathbb{R} \cup \{\infty\}$ are convex sets. A discussion of the topological nature of the sublevel sets needs the framework of lower semicontinuity.

Definition 3.5.7 An extended real-valued function f defined on a Hausdorff topological space X is called *lower semicontinuous* if

$$f(x) = \liminf_{y \to x} f(y) \quad \text{for all } x \in X.$$

In the same framework, a function g is called *upper semicontinuous* if $-g$ is lower semicontinuous.

The lower semicontinuous functions are precisely the functions for which all sublevel sets are closed (see Exercise 3). An important remark is that the supremum of any family of lower semicontinuous proper convex functions is a function of the same nature.

If the effective domain of a proper convex function is closed and f is continuous relative to $\operatorname{dom} f$, then f is lower semicontinuous. However, f can be lower semicontinuous without its effective domain being closed. The following function,

$$\varphi(x,y) = \begin{cases} y^2/2x & \text{if } x > 0, \\ \alpha & \text{if } x = y = 0, \\ \infty & \text{otherwise,} \end{cases}$$

is illustrative on what can happen at the boundary points of the effective domain. In fact, f is a proper convex function for each $\alpha \in [0, \infty]$. All points of its effective domain are points of continuity except the origin, where the limit does not exist. The function φ is lower semicontinuous for $\alpha = 0$.

The possibility of modifying the values of a proper convex function on the boundary of its effective domain to became lower semicontinuous is discussed in Exercises 2 and 3.

Exercises

1. Exhibit an example of a discontinuous linear functional defined on an infinite dimensional Banach space.

2. (W. Fenchel [79]) This exercise is devoted to an analogue of Proposition 1.3.4. Let f be a convex function on a convex subset U of $\mathbb{R}^n$. Prove:

(i) If x is a boundary point of U, then $\liminf_{y \to x} f(y) > -\infty$.

(ii) $\liminf_{y \to x} f(y) \leq f(x)$ if x is a boundary point of U that belongs to U.

(iii) Assume that U is open and consider the set V obtained from U by adding all the boundary points x for which $\liminf_{y \to x} f(y) < \infty$. Prove that V is convex and the function $g \colon V \to \mathbb{R}$ given by the formula

$$g(x) = \begin{cases} f(x) & \text{if } x \in U, \\ \liminf\limits_{y \to x} f(y) & \text{if } x \in V \cap \partial U, \end{cases}$$

is convex as well.

Remark. The last condition shows that every convex function can be modified at boundary points so that it becomes lower semicontinuous and convex.

3. Let f be an extended real-valued function defined on $\mathbb{R}^n$. Prove that the following conditions are equivalent:

(i) f is lower semicontinuous;
(ii) all sublevel sets $\{x|f(x) \leq \alpha\}$ are closed;
(iii) the epigraph $\mathrm{epi}(f)$ is a closed subset of $\mathbb{R}^{n+1}$;
(iv) if $x_m \to x$ in $\mathbb{R}^n$, then $f(x) \leq \liminf_{m\to\infty} f(x_m)$.

4. The *closure* of a proper convex function f is the function $\mathrm{cl}\, f$ whose epigraph is $\mathrm{epi}(\mathrm{cl}\, f) = \overline{\mathrm{epi}(f)}$. This function is lower semicontinuous and convex. Prove that $\mathrm{cl}\, f$ is the largest lower semicontinuous function minorizing f.

5. Let K be a polytope in $\mathbb{R}^n$. Prove that every bounded convex function on the relative interior of K has a unique extension to a continuous convex function on K.
Remark. D. Gale, V. Klee and R. T. Rockafellar [85] observed that this property characterizes the polytopes among the convex sets in $\mathbb{R}^n$.

3.6 Positively Homogeneous Functions

Many of the functions which arise naturally in convex analysis are real-valued functions f defined on a convex cone C in $\mathbb{R}^n$ (often $\mathbb{R}^n$ itself) that satisfy the relation

$$f(\lambda x) = \lambda f(x) \quad \text{for all } x \in C \text{ and all } \lambda \geq 0.$$

Such functions are called *positively homogeneous*. An important example is the norm mapping $\|\cdot\|$, which is defined on the whole space $\mathbb{R}^n$.

Lemma 3.6.1 *Let f be a positively homogeneous function defined on a convex cone C in $\mathbb{R}^n$. Then f is convex if and only if f is subadditive.*

Proof. Suppose that f is convex and $x, y \in C$. Then

$$\frac{1}{2} f(x+y) = f\Big(\frac{x+y}{2}\Big) \leq \frac{1}{2}\left(f(x) + f(y)\right)$$

and so $f(x+y) \leq f(x) + f(y)$.

Conversely, suppose that f is subadditive. Then

$$f((1-\lambda)x + \lambda y) \leq f((1-\lambda)x) + f(\lambda y) = (1-\lambda) f(x) + \lambda f(y)$$

for all $x, y \in C$ and $\lambda \in [0,1]$, which shows that f is convex. $\square$

Lemma 3.6.2 *Let f be a nonnegative positively homogeneous function defined on a convex cone C in $\mathbb{R}^n$ such that the sublevel set $\{x \in C \mid f(x) \leq 1\}$ is convex. Then f is a convex function.*

Proof. According to Lemma 3.6.1, it suffices to show that f is subadditive. For that, let $x, y \in C$ and choose scalars α and β such that $\alpha > f(x)$, and $\beta > f(y)$. Since f is nonnegative and positively homogeneous, $f(x/\alpha) \leq 1$ and $f(y/\beta) \leq 1$. Thus x/α and y/β both lie in the sublevel set of f at height 1. The assumed convexity of this sublevel set shows that

$$\frac{1}{\alpha+\beta} f(x+y) = f\Big(\frac{x+y}{\alpha+\beta}\Big) = f\Big(\frac{\alpha}{\alpha+\beta}\cdot\frac{x}{\alpha} + \frac{\beta}{\alpha+\beta}\cdot\frac{y}{\beta}\Big) \leq 1,$$

that is, $f(x+y) \leq \alpha+\beta$ whenever $\alpha > f(x)$, $\beta > f(y)$. Hence $f(x+y) \leq f(x) + f(y)$, which shows that f is subadditive. □

A sample of how the last lemma yields the convexity of some functions is as follows. Let $p \geq 1$ and consider the function f given on the nonnegative orthant $\mathbb{R}^n_+$ by the formula

$$f(x_1, \ldots, x_n) = (x_1^p + \cdots + x_n^p)^{1/p}.$$

Clearly, f is nonnegative and positively homogeneous, and f^p is convex as a sum of convex functions. Hence the sublevel set

$$\{x \in X \mid f(x) \leq 1\} = \{x \in X \mid f^p(x) \leq 1\}$$

is convex and this implies that f is a convex function. By Lemma 3.6.1 we conclude that f is subadditive, a fact which is equivalent with the Minkowski inequality.

In Section 1.8, we established Jensen's inequality in the context of finite measure spaces. Recently, P. Roselli and M. Willem [216] proved an extension of this inequality for all measure spaces, under the assumption that the convex function under attention is positively homogeneous and continuous. The basic ingredient in their proof is the following result, which is mostly a consequence of Theorem 1.5.2.

Lemma 3.6.3 *Suppose that $J\colon \mathbb{R}^2_+ \to \mathbb{R}$ is a positively homogeneous continuous function. Then the following assertions are equivalent:*

(i) *J is convex;*
(ii) *$\varphi = J(1, t)$ is convex;*
(iii) *there exists a subset $G \subset \mathbb{R}^2$ such that*

$$J(u, v) = \sup\{au + bv \mid (a, b) \in G\}.$$

Proof. Clearly, (i) $\Rightarrow$ (ii) and (iii) $\Rightarrow$ (i). For (ii) $\Rightarrow$ (iii) notice that $J(u, v) = uJ(1, v/u)$ if $u > 0$ and $J(u, v) = vJ(0, 1)$ if $u = 0$. Or, according to Theorem 1.5.2,

$$\varphi(t) = \sup\{a + bt \mid (a, b) \in G\}$$

where $G = \{(\varphi(s) - sb, b) \mid b \in \partial\varphi(s),\ s \in \mathbb{R}\}$. □

Theorem 3.6.4 (Roselli–Willem theorem) *Let $J\colon \mathbb{R}_+^2 \to \mathbb{R}$ be a positively homogeneous continuous convex function. Then for every measure space (X, Σ, μ) and every μ-integrable function $f\colon X \to \mathbb{R}_+^2$ for which $J \circ f$ is also μ-integrable, we have the inequality*

$$J\Big(\int_X f\,d\mu\Big) \leq \int_X J \circ f\,d\mu. \tag{3.6}$$

See Exercise 4 for a converse. Also, the role of $\mathbb{R}_+^2$ can be taken by every cone in $\mathbb{R}_+^n$.

Proof. Put $f = (f_1, f_2)$. According to Lemma 3.6.3, and Lebesgue's dominated convergence theorem,

$$\begin{aligned}\int_X (J \circ f)(x)\,d\mu &= \int_X \sup_{(a,b)\in G} (af_1 + bf_2)\,d\mu \\ &\geq \sup_{(a,b)\in G}\Big(a\int_X f_1\,d\mu + b\int_X f_2\,d\mu\Big) = J\Big(\int_X f\,d\mu\Big).\end{aligned}$$

□

The particular case where $f(x) = (|u(x)|^p, |v(x)|^p)$ and

$$J(u, v) = (u^{1/p} + v^{1/p})^p \quad (p \in \mathbb{R},\ p \neq 0)$$

gives us a very general version of Minkowski's inequality:

Theorem 3.6.5 *For $p \in (-\infty, 0) \cup [1, \infty)$ and $f, g \in L^p(\mu)$ we have*

$$\|f + g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p}, \tag{3.7}$$

while for $0 < p < 1$ the inequality works in the reverse sense,

$$\|f + g\|_{L^p} \geq \|f\|_{L^p} + \|g\|_{L^p}. \tag{3.8}$$

If f is not 0 almost everywhere, then we have equality if and only if $g = \lambda f$ almost everywhere, for some $\lambda \geq 0$.

Proof. In fact $J(1, t) = (1 + t^{1/p})^p$ is strictly convex for $0 < p < 1$ and strictly concave for $p \in (-\infty, 0) \cup (1, \infty)$. Then apply Theorem 3.6.4 above. □

There is a Minkowski type inequality even in the case $p = 0$. In fact, letting $p \to 0+$ in (3.8), and taking into account Exercise 1 (iv) in Section 1.8, we obtain the following:

Corollary 3.6.6 (Minkowski's inequality for $p = 0$) *Assume that (X, Σ, μ) is a finite measure space. For every $f, g \in L^1(\mu)$, $f, g \geq 0$ we have,*

$$\exp\left(\frac{1}{\mu(X)} \int_X \log(f(x) + g(x))\, d\mu\right) \geq \exp\left(\frac{1}{\mu(X)} \int_X \log f(x)\, d\mu\right) + \exp\left(\frac{1}{\mu(X)} \int_X \log g(x)\, d\mu\right).$$

For a direct proof of the discrete case see Exercise 4.

Another application of Theorem 3.6.4 is given by Hanner's inequalities. They improve Clarkson's inequalities and are the object of Exercise 7.

Exercises

1. (Support function) The *support function* of a nonempty compact convex set C in $\mathbb{R}^n$ is defined by

$$h(u) = \sup_{x \in C} \langle x, u \rangle, \quad u \in \mathbb{R}^n.$$

If $\|u\| = 1$, the set $H_\alpha = \{x \in \mathbb{R}^n \mid \langle x, u \rangle = \alpha\}$ describes a family of parallel hyperplanes, each having u as a normal vector; $\alpha = h(u)$ represents the value for which each H_α supports C and C is contained in the half-space H_α^-.

(i) Prove that the support function is positively homogeneous and convex.

(ii) Prove that $C = \{x \in \mathbb{R}^n \mid \langle x, u \rangle \leq h(u) \text{ for every } u \in \mathbb{R}^n\}$, which shows that C is the intersection of all half-spaces that contain it.

(iii) Conversely, let $h\colon \mathbb{R}^n \to \mathbb{R}$ be a positively homogeneous convex function. Prove that

$$C = \{x \in \mathbb{R}^n \mid \langle x, u \rangle \leq h(u) \text{ for every } u \in \mathbb{R}^n\}$$

is nonempty, compact, convex and its support function is h.

Remark. The notion of a support function can be attached to any nonempty convex set C in $\mathbb{R}^n$. See Section 3.7, Exercise 9.

2. *The Minkowski functional* (also called the *gauge function*) associated to a nonempty subset C of $\mathbb{R}^n$ is the function

$$p_C\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}, \quad p_C(x) = \inf\{\lambda > 0 \mid x \in \lambda C\}.$$

Here $\inf \emptyset = \infty$. Suppose that C is a closed convex set which contains the origin. Prove that:

(i) p_C is a positively homogeneous convex function.

(ii) The Minkowski functional of C is the support function of the polar set C°, and the Minkowski functional of C° is the support function of C.

(iii) C° is bounded if and only if $0 \in \operatorname{int} C$ (so by the bipolar theorem C is bounded if and only if $0 \in \operatorname{int} C^\circ$). Infer that the Minkowski functional of C is real-valued if $0 \in \operatorname{int} C$.

3. Find the support and the Minkowski functional of the following sets:

(i) $\{x \in \mathbb{R}^n \mid \sup\{|x_1|, \ldots, |x_n|\} \leq 1\}$;
(ii) $\{x \in \mathbb{R}^n \mid |x_1| + \cdots + |x_n| \leq 1\}$;
(iii) $\{x \in \mathbb{R}^n \mid x_1^2 + \cdots + x_n^2 \leq 1\}$.

4. Prove the following converse of Theorem 3.6.4: If J is continuous and the inequality (3.6) holds for every μ-integrable function $f\colon X \to \mathbb{R}_+^2$ for which $J \circ f$ is also μ-integrable, then J is positively homogeneous and convex.

5. (More on the Rogers–Hölder inequality) Extend the result of Section 1.2, Exercise 4, to the general context of measure spaces.

6. (Minkowski's inequality for $p = 0$: the discrete case) Prove that

$$\Big(\prod_{k=1}^{n}(x_k + y_k)\Big)^{1/n} \geq \Big(\prod_{k=1}^{n} x_k\Big)^{1/n} + \Big(\prod_{k=1}^{n} y_k\Big)^{1/n}$$

for every $x_1, \ldots, x_n, y_1, \ldots, y_n \geq 0$. When does equality occur?
[*Hint*: Use the following consequence of the AM–GM inequality:

$$\Big(\prod_{k=1}^{n} x_k\Big)^{1/n} = \inf\Big\{\frac{\alpha_1 x_1 + \cdots + \alpha_n x_n}{n} \;\Big|\; \alpha_1, \ldots, \alpha_n \geq 0, \prod_{k=1}^{n} \alpha_k = 1\Big\}. \quad]$$

7. (Hanner's inequalities) If $f, g \in L^p(\mu)$ and $2 \leq p < \infty$, then

$$\|f+g\|_{L^p}^p + \|f-g\|_{L^p}^p \leq (\|f\|_{L^p} + \|g\|_{L^p})^p + \big|\|f\|_{L^p} - \|g\|_{L^p}\big|^p,$$

equivalently (by making the replacements $f \to f+g$ and $g \to f-g$),

$$(\|f+g\|_{L^p} + \|f-g\|_{L^p})^p + \big|\|f+g\|_{L^p} - \|f-g\|_{L^p}\big|^p \geq 2^p(\|f\|_{L^p}^p + \|g\|_{L^p}^p).$$

If $1 < p \leq 2$, the above inequalities are reversed.
[*Hint*: Apply Theorem 3.6.4 for $f(x) = (|u(x)|^p, |v(x)|^p)$ and $J(u,v) = (u^{1/p} + v^{1/p})^p + |u^{1/p} - v^{1/p}|^p$.]

8. Prove that all nonempty closed convex subsets in a space $L^p(\mu)$ $(1 < p < \infty)$ are proximinal.
[*Hint*: Adapt the argument of Theorem 3.2.1, by using the inequalities of Hanner as a substitute for parallelogram's law.]

3.7 The Subdifferential

As already noted in Section 3.4, if f is a convex function (on an open convex subset U of a normed linear space E), then f has a supporting hyperplane at each point $a \in U$. This means the existence of a continuous linear functional h on E (the support of f at a) such that

$$f(x) \geq f(a) + h(x - a) \quad \text{for all } x \in U. \tag{3.9}$$

The set $\partial f(a)$ of all such functionals h constitutes the *subdifferential* of f at the point a.

By adapting the argument of Theorem 1.5.3 we can easily infer the following general result:

Theorem 3.7.1 *Suppose that U is an open convex set in a normed linear space E. Then a function $f\colon U \to \mathbb{R}$ is convex if and only if $\partial f(a) \neq \emptyset$ at all $a \in U$.*

When E is an $\mathbb{R}^n$ (or, more generally, a Hilbert space), all such h can be uniquely represented as

$$h(x) = \langle x, z \rangle \quad \text{for } x \in E.$$

In this case the inequality (3.9) becomes

$$f(x) \geq f(a) + \langle x - a, z \rangle \quad \text{for all } x \in U \tag{3.10}$$

and the subdifferential $\partial f(a)$ will be meant as the set of all such vectors z (usually called *subgradients*).

The analogue of Lemma 1.5.1 needs the notion of a directional derivative. Let f be a real-valued function defined on an open subset U of a Banach space E. The *one-sided directional derivatives of f at $a \in U$ relative to v* are defined to be the limits

$$f'_+(a\,; v) = \lim_{t \to 0+} \frac{f(a + tv) - f(a)}{t}$$

and

$$f'_-(a\,; v) = \lim_{t \to 0-} \frac{f(a + tv) - f(a)}{t}.$$

If both directional derivatives $f'_+(a\,; v)$ and $f'_-(a\,; v)$ exist and they are equal, we shall call their common value the *directional derivative of f at a, relative to v* (also denoted $f'(a\,; v)$). Notice that the one-sided directional derivatives are positively homogeneous and subadditive (as a function of v), see Exercise 1. Taking into account the formula, $f'_+(a\,; v) = -f'_-(a\,; -v)$, we infer that the directional derivatives (when they exist) are linear.

The directional derivatives relative to the vectors of the canonical basis of $\mathbb{R}^n$ are nothing but the partial derivatives.

If f is convex, then for each pair $(a,v) \in U \times E$ there exists an interval $(-\varepsilon,\varepsilon)$ on which the function $t \to f(a+tv)$ is well-defined and convex. Taking into account Theorem 1.3.3, it follows that every convex function admits one-sided directional derivatives at any point and that

$$f'_+(a\,;v) \geq f'_-(a\,;v).$$

As $f'_-(a\,;v) = -f'_+(a\,;-v)$, the above discussion yields the following analogue of Lemma 1.5.1:

Lemma 3.7.2 *Suppose that f is a convex function defined on an open convex subset U of $\mathbb{R}^n$. Then $z \in \partial f(a)$ if and only if $f'_+(a\,;v) \geq \langle z, v\rangle$ for all $v \in \mathbb{R}^n$.*

In the finite dimensional case, $\partial f(a)$ is a singleton precisely when f has a directional derivative $f'(a\,;v)$ relative to any v. In that case, $\partial f(a)$ consists of the mapping $v \to f'(a\,;v)$. See Theorem 3.8.2.

If $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a lower semicontinuous proper convex function, we say that $z \in \mathbb{R}^n$ is a *subgradient* of f at $a \in \operatorname{dom} f$ if

$$f(x) \geq f(a) + \langle x-a, z\rangle \quad \text{for all } x \in \mathbb{R}^n. \tag{3.11}$$

We call the set $\partial f(a)$, of all subgradients of f at a, the *subdifferential* of f (at the point a).

A derivative is a local property, while the subgradient definition (3.10) describes a global property. An illustration of this idea is the following remark: *for any lower semicontinuous proper function $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, the point a is a global minimizer of f if and only if*

$$0 \in \partial f(a).$$

The subdifferential calculus is presented in Exercises 3–5.

The *subdifferential* of f is defined as the set-valued map ∂f which associates to each $x \in \mathbb{R}^n$ the subset $\partial f(x) \subset \mathbb{R}^n$. Equivalently, ∂f may be seen as a graph in $\mathbb{R}^n \times \mathbb{R}^n$. Given two set-valued maps $u, v\colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$, we define

- *domain of* u, $\operatorname{dom} u = \{x \mid u(x) \neq \emptyset\}$;
- *graph of* u, $\operatorname{graph} u = \{(x,y) \mid y \in u(x)\}$;
- *inverse of* u, $u^{-1}(y) = \{x \mid y \in u(x)\}$;
- $u \subset v$, if the graph of u is contained in the graph of v.

Definition 3.7.3 A set-valued map $u\colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ is said to be *monotone* if it verifies

$$\langle x_1 - x_2, y_1 - y_2\rangle \geq 0$$

for all $x_1, x_2 \in \mathbb{R}^n$ and all $y_1 \in u(x_1)$, $y_2 \in u(x_2)$. A monotone function u is called *maximal monotone* when it is maximal with respect to inclusion in the class of monotone functions, that is, if the following implication holds:

$$v \supset u \text{ and } v \text{ monotone} \implies v = u.$$

According to Zorn's lemma, for each monotone function u there exists a maximal monotone function $\widetilde{u}$ which includes u.

The graph of any maximal monotone map $u\colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ is closed and thus it verifies the following conditions of upper semicontinuity:

$$x_k \to x,\ y_k \to y, \text{ and } y_k \in u(x_k) \text{ for all } k \in \mathbb{N} \implies y \in u(x).$$

We shall prove the existence of a one-to-one correspondence between graphs of maximal monotone maps and graphs of nonexpansive functions. Recall that a function $h\colon \mathbb{R}^n \to \mathbb{R}^n$ is called *nonexpansive* if its Lipschitz constant verifies

$$\operatorname{Lip}(h) = \sup_{x \neq y} \frac{\|h(x) - h(y)\|}{\|x - y\|} \leq 1.$$

We shall need the following result concerning the extension of Lipschitz functions:

Theorem 3.7.4 (M. D. Kirszbraun) *Suppose that A is a subset of $\mathbb{R}^n$ and $f\colon A \to \mathbb{R}^m$ is a Lipschitz function. Then there exists a Lipschitz function $\widetilde{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ such that $\widetilde{f} = f$ on A and $\operatorname{Lip}(\widetilde{f}) = \operatorname{Lip}(f)$. Moreover, we may choose $\widetilde{f}$ convex, when A and f are also convex.*

Proof. When $m = 1$, we may choose

$$\widetilde{f}(x) = \inf_{y \in A} \{f(y) + \operatorname{Lip}(f) \cdot \|x - y\|\}.$$

In the general case, a direct application of this remark at the level of components of f leads to an extension $\widetilde{f}$ with $\operatorname{Lip}(\widetilde{f}) \leq \sqrt{m}\operatorname{Lip}(f)$. The existence of an extension with the same Lipschitz constant is described in [77, Section 2.10.43, p. 201]. □

The aforementioned correspondence between graphs is realized by the *Cayley transform*, that is, by the linear isometry

$$\Phi\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n, \quad \Phi(x, y) = \frac{1}{\sqrt{2}}(x + y, -x + y).$$

When $n = 1$, the Cayley transform represents a clockwise rotation of angle $\pi/4$. The precise statement of this correspondence is as follows:

Theorem 3.7.5 (G. Minty [165]) *Let $u\colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ be a maximal monotone map. Then $J = (I + u)^{-1}$ is defined on the whole $\mathbb{R}^n$ and $\Phi(\operatorname{graph} u)$ is the graph of a nonexpansive function $v\colon \mathbb{R}^n \to \mathbb{R}^n$, given by*

$$v(x) = x - \sqrt{2}\,(I + u)^{-1}(\sqrt{2}\,x). \tag{3.12}$$

Conversely, if $v\colon \mathbb{R}^n \to \mathbb{R}^n$ is a nonexpansive function, then the inverse image of $\operatorname{graph} v$ under Φ is the graph of a maximal monotone function on $\mathbb{R}^n$.

Here I denotes the identity map of $\mathbb{R}^n$.

Proof. Let u be a monotone map and let v be the set-valued function whose graph is $\Phi(\operatorname{graph} u)$. We shall show that v is nonexpansive in its domain (and thus single-valued). In fact, given $x \in \mathbb{R}^n$, we have

$$y \in v(x) \quad \text{if and only if} \quad \frac{x+y}{\sqrt{2}} \in u\Big(\frac{x-y}{\sqrt{2}}\Big) \tag{3.13}$$

and this yields $y \in x - \sqrt{2}\,(I+u)^{-1}(\sqrt{2}x)$ for all $y \in v(x)$.

Now, if $x_k \in \mathbb{R}^n$ and $y_k \in v(x_k)$ for $k = 1, 2$, we infer from (3.13) that

$$\langle (x_1 - y_1) - (x_2 - y_2), (x_1 + y_1) - (x_2 + y_2) \rangle \geq 0,$$

hence $\|y_1 - y_2\|^2 \leq \|x_1 - x_2\|^2$. This shows that v is indeed nonexpansive.

The same argument shows that Φ^{-1} maps graphs of nonexpansive functions into graphs of monotone functions.

Assuming that u is maximal monotone, we shall show that the domain of v is $\mathbb{R}^n$. In fact, if the contrary were true, we could apply Theorem 3.7.4 to extend v to a nonexpansive function $\widetilde{v}$ defined on the whole $\mathbb{R}^n$, and then $\Phi^{-1}(\operatorname{graph} \widetilde{v})$ provides a monotone extension of u, which contradicts the maximality of u. □

Corollary 3.7.6 *Let $u\colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ be a maximal monotone map. Then $J = (I+u)^{-1}$ is a nonexpansive map of $\mathbb{R}^n$ into itself.*

Proof. It is easy to see that $I + u$ (and thus $(I + u)^{-1}$) is monotone. By Theorem 3.7.5, the maximality of u yields the surjectivity of $I + u$, hence $\operatorname{dom}(I + u)^{-1} = \mathbb{R}^n$. In order to prove that $(I + u)^{-1}$ is also a nonexpansive function, let us consider points $x_k \in \mathbb{R}^n$ and $y_k \in u(x_k)$ (for $k = 1, 2$). Then

$$\begin{aligned} \|x_1 - x_2\|^2 &\leq \langle x_1 - x_2, x_1 - x_2 + y_1 - y_2 \rangle \\ &\leq \|x_1 - x_2\| \cdot \|x_1 + y_1 - (x_2 + y_2)\|, \end{aligned} \tag{3.14}$$

which yields $\|x_1 - x_2\| \leq \|(x_1+y_1)-(x_2+y_2)\|$. Particularly, if $x_1+y_1 = x_2+y_2$, then $x_1 = x_2$, and this shows that $(I + u)^{-1}$ is single-valued. Consequently, $(I+u)^{-1}(x_k+y_k) = x_k$ for $k = 1, 2$ and thus (3.14) yields the nonexpansivity of $(I + u)^{-1}$. □

An important class of maximal monotone maps is provided by the subdifferentials of convex functions.

Theorem 3.7.7 *If $f\colon \mathbb{R}^n \to \mathbb{R}\cup\{\infty\}$ is a lower semicontinuous proper convex function, then ∂f is a maximal monotone function such that*

$$\operatorname{int}\operatorname{dom} f \subset \operatorname{dom} \partial f \subset \operatorname{dom} f.$$

Proof. The fact that ∂f is monotone follows from (3.11). According to Theorem 3.7.5, the maximality of ∂f is equivalent to the surjectivity of $\partial f + I$. To prove that $\partial f + I$ is onto, let us fix arbitrarily $y \in \mathbb{R}^n$, and choose $x \in \mathbb{R}^n$ as the unique minimizer of the coercive lower semicontinuous function

$$g\colon x \to f(x) + \frac{1}{2}\,\|x\|^2 - \langle x, y\rangle.$$

Then $0 \in \partial g(x)$, which yields $y \in \partial(f(x) + \|x\|^2/2) = (\partial f + I)(x)$. $\square$

According to W. Fenchel [78], the *conjugate* (or the *Legendre transform*) of a function $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is the function $f^*\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ defined by

$$f^*(y) = \sup_{x\in\mathbb{R}^n} [\langle x, y\rangle - f(x)].$$

The function f^* is always lower semicontinuous and convex, and, if the effective domain of f is nonempty, then f^* never takes the value $-\infty$. Clearly, $f \le g$ yields $f^* \ge g^*$ (and thus $f^{**} \le g^{**}$). Also, the following generalization of *Young's inequality* holds true: If f is a proper convex function then so is f^* and

$$f(x) + f^*(y) \ge \langle x, y\rangle \quad \text{for all } x, y \in \mathbb{R}^n.$$

Equality holds if and only if $\langle x, y\rangle \ge f(x) + f^*(y)$, equivalently, when $f(z) \ge f(x) + \langle y, z - x\rangle$ for all z (that is, when $y \in \partial f(x)$).

By Young's inequality we infer that

$$f(x) \ge \sup_y [\langle x, y\rangle - f^*(y)] = f^{**}(x) \quad \text{for all } x \in \mathbb{R}^n.$$

All the material in Section 1.7 on conjugate functions can be adapted *mutatis mutandis* to the context of several variables. See [213] for details. We can prove that conjugacy induces a bijection between lower semicontinuous proper convex functions.

Theorem 3.7.8 *Suppose that $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a proper convex function. Then the following assertions are equivalent:*

(i) *f is lower semicontinuous*;
(ii) $f = f^{**}$;
(iii) *f is the pointwise supremum of the family of all affine functions h such that $h \le f$.*

Proof. Clearly, (ii) $\Rightarrow$ (i) and (ii) $\Rightarrow$ (iii). Since any affine minorant h of f verifies $h = h^{**} \le f^{**} \le f$, it follows that (iii) $\Rightarrow$ (ii). The implication (i) $\Rightarrow$ (iii) can be proved easily by using the basic separation theorem. See [38, pp. 76–77].

Alternatively, we can show that (i) $\Rightarrow$ (ii). If $x \in \operatorname{int}(\operatorname{dom} f)$, then $\partial f(x)$ is nonempty and for each $y \in \partial f(x)$ we have $\langle x, y\rangle = f(x) + f^*(y)$, hence $f(x) = \langle x, y\rangle - f^*(y) \le f^{**}(x)$. In the general case, we may use an approximation argument. See Section 3.8, Exercise 7. $\square$

Remark 3.7.9 Conjugacy offers a convenient way to recognize properties like coercivity and convexity. Here are the precise statements.

(i) (J.-J. Moreau and R. T. Rockafellar, see [38, p. 78]) *A lower semicontinuous proper convex function has bounded level subsets if and only if its conjugate is continuous at the origin.*

(ii) (J.-B. Hiriart-Urruty [104]) *Suppose that* $f\colon \mathbb{R}^n \to \mathbb{R}\cup\{\infty\}$ *is lower semicontinuous, not identically* ∞ *and* $\lim_{x\to\infty} f(x)/\|x\| = \infty$. *If the conjugate of* f *is differentiable, then* f *is necessarily convex.*

Exercises

1. (Subadditivity of the directional derivatives) Suppose that f is a convex function (on an open convex set U in a normed linear space E). For $a \in U$, $u, v \in E$ and $t > 0$ small enough, show that

$$\frac{f(a+t(u+v)) - f(a)}{t} \leq \frac{f(a+2tu) - f(a)}{2t} + \frac{f(a+2tv) - f(a)}{2t}$$

and conclude that $f'_+(a\,;u+v) \leq f'_+(a\,;u) + f'_+(a\,;v)$.

2. Compute $\partial f(0)$ when $f(x) = \|x\|$ is the Euclidean norm on $\mathbb{R}^n$.

3. Suppose that f, f_1, f_2 are convex functions on $\mathbb{R}^n$ and $a \in \mathbb{R}^n$.

(i) Infer from Lemma 3.7.2 that

$$f'_+(a;v) = \sup\{\langle z, v\rangle \mid z \in \partial f(a)\} \quad \text{for all } v \in \mathbb{R}^n.$$

(ii) Let λ_1 and λ_2 be two positive numbers. Prove that

$$\partial(\lambda_1 f_1 + \lambda_2 f_2)(a) = \lambda_1 \partial f_1(a) + \lambda_2 \partial f_2(a).$$

Remark. In the general setting of proper convex functions, only the inclusion $\supset$ works. The equality needs additional assumptions, for example, the existence of a common point in the convex sets $\operatorname{ri}(\operatorname{dom} f_k)$ for $k = 1, \ldots, m$. See [213, p. 223].

4. Let f be a proper convex function on $\mathbb{R}^n$ and let A be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$. Prove the formula

$$\partial(f\circ A)(x) \supset A^*\partial f(Ax).$$

Remark. The equality needs additional assumptions. For example, it works when the range of A contains a point of $\operatorname{ri}(\operatorname{dom} f_k)$. See [213, p. 225].

5. (Subdifferential of a max-function) Suppose that $f_1, \ldots, f_m$ are convex functions on $\mathbb{R}^n$ and set $f = \max\{f_1, \ldots, f_m\}$. For $a \in \mathbb{R}^n$ set

$$J(a) = \{j \mid f_j(a) = f(a)\}.$$

Prove that $\partial f(a) = \operatorname{co}\{\partial f_j(a) \mid j \in J(a)\}$.

6. Show, by examples, that the two inclusions in Theorem 3.7.7 may be strict, and dom ∂f may not be convex.

7. (R. T. Rockafellar [213, pp. 238–239]) A *cyclically monotone* map is any set-valued function $u\colon \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ such that

$$\langle x_2 - x_1, y_1\rangle + \langle x_3 - x_2, y_2\rangle + \cdots + \langle x_1 - x_m, y_m\rangle \leq 0$$

for all finite families of points $(x_k, y_k) \in \mathbb{R}^n \times \mathbb{R}^n$ with $y_k \in u(x_k)$, $k \in \{1, \ldots, m\}$. By the inequality (3.11), the subdifferential of any lower semicontinuous proper convex function on $\mathbb{R}^n$ is cyclically monotone. Prove the following integrability result: for any cyclically monotone map there exists a lower semicontinuous proper convex function f such that $u \subset \partial f$. [*Hint*: Consider the function

$$f(x) = \sup\{\langle x - x_m, y_m\rangle + \langle x - x_{m-1}, y_{m-1}\rangle + \cdots + \langle x - x_0, y_0\rangle\},$$

where the supremum is taken over all finite sets of pairs $(x_k, y_k) \in \mathbb{R}^n \times \mathbb{R}^n$ such that $y_k \in u(x_k)$ for all k.]

8. Suppose that f is a convex function on $\mathbb{R}^n$. Prove that $f = f^*$ if and only if $f(x) = \|x\|^2/2$.

9. (Support function) The notion of a *support function* can be attached to any nonempty convex set C in $\mathbb{R}^n$, by defining it as the conjugate of the indicator function of C. Prove that the support function of $C = \{(x, y) \in \mathbb{R}^2 \mid x + y^2/2 \leq 0\}$ is $\delta_C^*(x, y) = y^2/2x$ if $x > 0$, $\delta_C^*(0, 0) = 0$ and $\delta_C^*(x, y) = \infty$ otherwise. Infer that δ_C^* is a lower semicontinuous proper convex function.

10. Calculate the support function for the set

$$C = \{A \in \mathrm{Sym}^{++}(n, \mathbb{R}) \mid \mathrm{trace}(A) = 1\}.$$

11. (Legendre transform) Let $f : \mathbb{R}^n \to \mathbb{R}$ be a strictly convex function of class C^1, such that $f(x)/\|x\| \to \infty$ as $\|x\| \to \infty$. Prove that:

(i) The map $x \to \nabla f(x)$ is a homeomorphism (from $\mathbb{R}^n$ onto itself);
(ii) $f^\star(y) = \langle y, (\nabla f)^{-1} y\rangle - f((\nabla f)^{-1} y)$ for all $y \in \mathbb{R}^n$;
(iii) $f^\star$ is a C^1 function and $\nabla f^\star = (\nabla f)^{-1}$.

[*Hint*: For every $x, y \in \mathbb{R}^n$, $x, y \neq 0$, the function $g(t) = f(x+ty)$ is strictly convex on $\mathbb{R}$ and thus $g'(1) - g'(0) = \langle \nabla f(x+y) - \nabla f(y), y\rangle > 0$. This shows that ∇f is one-to-one. Let $z \in \mathbb{R}^n$. Since $g(x) = f(x) - \langle x, z\rangle$ is coercive and C^1, it attains a global minimum at a point a for which $\nabla g(a) = \nabla f(a) - z = 0$. Hence ∇f is onto. The inequality $f(0) + \langle \nabla f(x), x\rangle \geq f(x)$ yields $\|\nabla f(x)\| \to \infty$ as $\|x\| \to \infty$. Therefore the inverse image under ∇f of every compact set is compact too, a fact which assures the continuity of $(\nabla f)^{-1}$.]

3.8 Differentiability of Convex Functions

The problem of differentiability of a convex function defined on an open subset U of a Banach space E can be treated in the setting of Fréchet differentiability or in the more general setting of Gâteaux differentiability.

The *Fréchet differentiability* (or, simply, the *differentiability*) of f at a point a means the existence of a continuous linear functional $df(a)\colon E \to \mathbb{R}$ such that

$$\lim_{x\to a} \frac{|f(x) - f(a) - df(a)(x-a)|}{\|x-a\|} = 0.$$

Equivalently,

$$f(x) = f(a) + df(a)(x-a) + \omega(x)\|x-a\| \quad \text{for } x \in U,$$

where $\omega\colon U \to \mathbb{R}$ is a function such that $\omega(a) = \lim_{x\to a}\omega(x) = 0$. When $E = \mathbb{R}^n$, the functional $df(a)$ can be computed via the formula

$$df(a)(v) = \langle \nabla f(a), v\rangle,$$

where

$$\nabla f(a) = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(a)\mathrm{e}_k$$

represents the *gradient of f at a*.

A function $f\colon U \to \mathbb{R}$ is said to be *Gâteaux differentiable* at a point a if the directional derivative $f'(a\,;v)$ exists for every $v \in E$ and defines a continuous linear functional $f'(a)\colon v \to f'(a\,;v)$ on E. It is straightforward that differentiability implies Gâteaux differentiability and also the equality

$$f'(a) = df(a).$$

For convex functions on open subsets of $\mathbb{R}^n$, Gâteaux and Fréchet differentiability agree:

Theorem 3.8.1 *Suppose that a convex function f defined on an open convex set U in $\mathbb{R}^n$ possesses all its partial derivatives $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}$ at some point $a \in U$. Then f is differentiable at a.*

Proof. Since U is open, there is a $r > 0$ such that $B_r(a) \subset U$. We have to prove that the function

$$g(u) = f(a+u) - f(a) - \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(a)u_k$$

defined for all $u = (u_1, \ldots, u_n)$ with $\|u\| < r$, verifies $\lim_{\|u\|\to 0} g(u)/\|u\| = 0$.

Clearly, the function g is convex. Then

$$0 = g(0) = g\Big(\frac{u + (-u)}{2}\Big) \leq \frac{1}{2}\,(g(u) + g(-u)),$$

which yields $g(u) \geq -g(-u)$. On the other hand, for each u with $n\|u\| < r$, we have

$$\begin{aligned} g(u) &= g\Big(\frac{1}{n}\sum_{k=1}^{n} nu_k e_k\Big) \leq \frac{1}{n}\sum_{k=1}^{n} g(nu_k e_k) \\ &= \sum_{\{k\,|\,u_k \neq 0\}} u_k\,\frac{g(nu_k e_k)}{nu_k} \leq \|u\| \sum_{\{k\,|\,u_k \neq 0\}} \Big|\frac{g(nu_k e_k)}{nu_k}\Big|. \end{aligned}$$

Similarly,

$$g(-u) \leq \|u\| \sum_{\{k\,|\,u_k \neq 0\}} \Big|\frac{g(-nu_k e_k)}{nu_k}\Big|.$$

Then

$$-\|u\| \sum_{\{k\,|\,u_k \neq 0\}} \Big|\frac{g(-nu_k e_k)}{nu_k}\Big| \leq -g(-u) \leq g(u) \leq \|u\| \sum_{\{k\,|\,u_k \neq 0\}} \Big|\frac{g(nu_k e_k)}{nu_k}\Big|$$

and it remains to remark that $g(nu_k e_k)/(nu_k) \to 0$ as $u_k \to 0$. $\square$

The condition of differentiability is equivalent to the uniqueness of the support function:

Theorem 3.8.2 *Let f be a convex function defined on an open convex set U in $\mathbb{R}^n$. Then f is differentiable at a if and only if f has a unique support at a.*

Proof. Suppose that $f'(a\,;v)$ exists for every v. If $h\colon E \to \mathbb{R}$ is a support of f at a, then

$$f(a + \varepsilon v) - f(a) \geq \varepsilon h(v)$$

for sufficiently small $\varepsilon > 0$, which yields $f'(a\,;v) \geq h(v)$. Replacing v by $-v$, and taking into account that the directional derivative is linear in v, we obtain

$$-f'(a\,;v) = f'(a\,;-v) \geq -h(v)$$

from which we conclude that $h(v) = f'(a\,;v)$.

Suppose now that f has a unique support h at a and choose a number λ such that

$$f'_{-}(a\,;e_1) \leq \lambda \leq f'_{+}(a\,;e_1).$$

Then the line L in $\mathbb{R}^{n+1}$ given by $t \to (a+te_1, f(a)+\lambda t)$ meets the epigraph of f at $(a, f(a))$. Since $f(a + te_1) \geq f(a) + \lambda t$ as long as $a + te_1 \in U$, the line L does not meet the interior of the epigraph of f. By the Hahn–Banach theorem (see Appendix A) we infer the existence of a supporting hyperplane to the epigraph of f at $(a, f(a))$ which contains L. The uniqueness of the support of f at a shows that this hyperplane must be the graph of h. Since

$$h(a+te_1) = f(a) + \lambda t = h(a) + \lambda t$$

for all $t \in \mathbb{R}$, it follows that only one λ can be found satisfying the above choice, and thus $f'_-(a\,; e_1) = f'_+(a\,; e_1)$. In other words we established the existence of $\partial f/\partial x_1$ at a. Similarly, one can prove the existence of all partial derivatives at a so, by Theorem 3.8.1, the function f is differentiable at a. □

In the context of several variables, the set of points where a convex function is not differentiable can be uncountable, though still negligible:

Theorem 3.8.3 *Suppose that f is a convex function on an open subset U of $\mathbb{R}^n$. Then f is differentiable almost everywhere in U.*

Proof. Consider first the case when U is also bounded. According to Theorem 3.8.1 we must show that each of the sets

$$E_k = \left\{ x \in U \;\middle|\; \frac{\partial f}{\partial x_k}(x) \text{ does not exist} \right\}$$

is Lebesgue negligible. The measurability of E_k is a consequence of the fact that the limit of a pointwise converging sequence of measurable functions is measurable too. In fact, the formula

$$f'_+(x, e_k) = \lim_{j\to\infty} \frac{f(x+e_k/j) - f(x)}{1/j}$$

motivates the measurability of one-sided directional derivative $f'_+(x, e_k)$ and a similar argument applies for $f'_-(x, e_k)$. Consequently the set

$$E_k = \{x \in U \mid f'_+(x, e_k) - f'_-(x, e_k) > 0\}$$

is measurable. Being bounded, it is also integrable. By Fubini's theorem,

$$\begin{aligned} m(E_k) &= \int_{\mathbb{R}^n} \chi_{E_k}\, dx \\ &= \int_{\mathbb{R}} \cdots \left(\int_{\mathbb{R}} \chi_{E_k}\, dx_i \right) dx_1 \cdots dx_{i-1}\, dx_{i+1} \cdots dx_n \end{aligned}$$

and the interior integral is zero since f is convex as a function of x_i (and thus differentiable except at an enumerable set of points).

If U is arbitrary, the argument above shows that all the sets $E_k \cap B_n(0)$ are negligible. Or, $E_k = \bigcup_{n=1}^{\infty}(E_k \cap B_n(0))$ and a countable union of negligible sets is negligible too. □

The function $f(x,y) = \sup\{x, 0\}$ is convex on $\mathbb{R}^2$ and nondifferentiable at the points of y-axis (which constitutes an uncountable set).

The coincidence of Gâteaux and Fréchet differentiability is no longer true in the context of infinite dimensional spaces.

Theorem 3.8.4 *Let E be a Banach space such that for each continuous convex function $f\colon E \to \mathbb{R}$, every point of Gâteaux differentiability is also a point of Fréchet differentiability. Then E is finite dimensional.*

The proof we present here is due to J. M. Borwein and A. S. Lewis [38], and depends on a deep result in Banach space theory:

Theorem 3.8.5 (The Josephson–Nissenzweig theorem[117], [190]) *If E is a Banach space such that*

$$x_n' \to 0 \textit{ in the weak-star topology of } E' \quad \textit{implies} \quad \|x_n'\| \to 0,$$

then E is finite dimensional.

Recall that the weak-star topology of E' is the topology of pointwise convergence.

Proof of Theorem 3.8.4. Consider a sequence $(x_n')_n$ of norm-1 functionals in E' and a sequence $(\alpha_n)_n$ of real numbers such that $\alpha_n \downarrow 0$. Then the function

$$f(x) = \sup_n [\langle x, x_n' \rangle - \alpha_n]$$

is convex and continuous and, moreover,

$$f \text{ is Gâteaux differentiable at } 0 \iff x_n'(x) \to 0 \text{ for all } x \in E$$
$$f \text{ is Fréchet differentiable at } 0 \iff \|x_n'\| \to 0.$$

The proof ends by applying the Josephson–Nissenzweig theorem. ☐

Convolution by smooth functions provides us with a powerful technique for approximating locally integrable functions by C^∞ functions. Particularly, this applies to the convex functions.

Let φ be a *mollifier*, that is, a nonnegative function in $C_c^\infty(\mathbb{R}^n)$ such that

$$\int_{\mathbb{R}^n} \varphi \, dx = 1 \quad \text{and} \quad \operatorname{supp} \varphi \subset \overline{B}_1(0).$$

The standard example of such a function is given by

$$\varphi(x) = \begin{cases} C \exp(-1/(1 - \|x\|^2)) & \text{if } \|x\| < 1, \\ 0 & \text{if } \|x\| \geq 1, \end{cases}$$

where C is chosen such that $\int_{\mathbb{R}^n} \varphi \, dx = 1$. Each mollifier φ gives rise to an one-parameter family of nonnegative functions

$$\varphi_\varepsilon(x) = \frac{1}{\varepsilon^n} \varphi\Big(\frac{x}{\varepsilon}\Big), \quad \varepsilon > 0$$

with similar properties:

$$\varphi_\varepsilon \in C_c^\infty(\mathbb{R}^n), \quad \operatorname{supp} \varphi_\varepsilon \subset \overline{B}_\varepsilon(0) \quad \text{and} \quad \int_{\mathbb{R}^n} \varphi_\varepsilon \, dx = 1.$$

The following lemma is standard and available in many places. For example, see [74, pp. 122–125] or [252, pp. 22–23].

Lemma 3.8.6 *Suppose that $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$ and $(\varphi_\varepsilon)_{\varepsilon>0}$ is the one-parameter family of functions associated to a mollifier φ. Then:*

(i) *the functions*

$$f_\varepsilon = \varphi_\varepsilon * f$$

belong to $C^\infty(\mathbb{R}^n)$ and

$$D^\alpha f_\varepsilon = D^\alpha \varphi_\varepsilon * f$$

for every multi-index α;

(ii) *$f_\varepsilon(x) \to f(x)$ whenever x is a point of continuity of f. If f is continuous on an open subset U, then f_ε converges uniformly to f on each compact subset of U;*

(iii) *if $f \in L^p(\mathbb{R}^n)$ (for some $p \in [1, \infty)$), then $f_\varepsilon \in L^p(\mathbb{R}^n)$, $\|f_\varepsilon\|_{L^p} \leq \|f\|_{L^p}$ and $\lim_{\varepsilon \to 0} \|f_\varepsilon - f\|_{L^p} = 0$;*

(iv) *if f is a convex function on an open convex subset U of $\mathbb{R}^n$, then f_ε is convex too.*

An application of Lemma 3.8.6 is given in Exercise 5.

A nonlinear analogue of mollification is offered by the *infimal convolution*, which for two proper convex functions $f, g \colon E \to \mathbb{R} \cup \{\infty\}$ is defined by the formula

$$(f \odot g)(x) = \inf\{f(x-y) + g(y) \mid y \in E\};$$

the value $-\infty$ is allowed. If $(f \odot g)(x) > -\infty$ for all x, then $f \odot g$ is a proper convex function. For example, this happens when both functions f and g are nonnegative (or, more generally, when there exists an affine function $h \colon E \to \mathbb{R}$ such that $f \geq h$ and $g \geq h$).

By computing the infimal convolution of the norm function and the indicator function of a convex set C, we get

$$(\|\cdot\| \odot \delta_C)(x) = \inf_{y \in C} \|x - y\| = d_C(x),$$

a fact which implies the convexity of the distance function.

A standard way to approximate from below a lower semicontinuous proper convex function $f \colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is the *Moreau–Yosida approximation*:

$$\begin{aligned} f_\varepsilon(x) &= \Big(f \odot \frac{1}{2\varepsilon}\|\cdot\|^2\Big)(x) \\ &= \inf_{y \in \mathbb{R}^n} \Big\{ f(y) + \frac{1}{2\varepsilon}\|x-y\|^2 \Big\} \end{aligned}$$

for $x \in \mathbb{R}^n$ and $\varepsilon > 0$. The functions f_ε are well-defined and finite for all x because the function $y \to f(y) + \frac{1}{2\varepsilon}\|x-y\|^2$ is lower semicontinuous and also coercive (due to the existence of a support for f).

Lemma 3.8.7 *The Moreau–Yosida approximates f_ε are differentiable convex functions on $\mathbb{R}^n$ and $f_\varepsilon \to f$ as $\varepsilon \to 0$. Moreover, $\partial f_\varepsilon = (\varepsilon I + (\partial f)^{-1})^{-1}$ as set-valued maps.*

The first statement is straightforward. The proof of the second one may be found in [4], [14] and [43].

As J. M. Lasry and P.-L. Lions [139] observed, the infimal convolution provides an efficient regularization procedure for (even degenerate) elliptic equations. This explains the *Lax formula*,

$$u(x,t) = \sup_{y \in \mathbb{R}^n} \left\{ v(y) - \frac{1}{2t}\|x-y\|^2 \right\},$$

for the solution of the *Hamilton-Jacobi equation*,

$$\frac{\partial u}{\partial t} - \frac{1}{2}\|\nabla u\|^2 = 0 \quad \text{for } x \in \mathbb{R}^n,\ t > 0$$
$$u|_{t=0} = v \quad \text{on } \mathbb{R}^n.$$

Exercises

1. Prove that the norm function on $C([0,1])$,

$$\|x\| = \sup\{|x(t)| \mid t \in [0,1]\},$$

is not differentiable at any point, but it is Gâteaux differentiable at those x of the unit sphere for which $|x(t_0)| = 1$ is attained for only one value t_0.

2. Prove that the norm function of a Hilbert space is differentiable at any point $x \neq 0$.

3. Let C be a nonempty closed subset of $\mathbb{R}^n$ and let

$$d_C \colon \mathbb{R}^n \to \mathbb{R}, \quad d_C(x) = \inf\{\|x-y\| \mid y \in C\}$$

be the distance function.

(i) Prove that $\varphi = d_C^2$ verifies the relation

$$\varphi(x+y) = \varphi(x) + \varphi'(x;y) + \varepsilon(y)\|y\|,$$

where $\varphi'(x;y) = \min\{\langle 2y, x-z\rangle \mid z \in \mathcal{P}_C(x)\}$ and $\lim_{y\to 0}\varepsilon(y) = \varepsilon(0) = 0$ (and thus φ is Gâteaux differentiable everywhere).

(ii) Suppose that C is also convex. Infer the formula

$$\nabla \frac{d_C^2}{2}(x) = x - P_C(x) \quad \text{for all } x \in \mathbb{R}^n.$$

(iii) Prove that d_C is differentiable at a point $x \in \Omega = \mathbb{R}^n \backslash C$ if and only if $\mathcal{P}_C(x)$ is a singleton.

(iv) Consider the function $f_C(x) = \|x\|^2/2$ if $x \in C$, and $f_C(x) = \infty$ if $x \in \mathbb{R}^n \backslash C$ (where C is a nonempty closed subset of $\mathbb{R}^n$). Notice that $f_C^*(y) = [\|y\|^2 - d_C^2(y)]/2$ and infer from Remark 3.7.9 (ii) the conclusion of Bunt's theorem (Theorem 3.2.2 above).

4. Infer from the Josephson–Nissenzweig theorem that any Banach space E for which all continuous convex functions on E are bounded on bounded subsets is finite dimensional.
[*Hint*: Consider a sequence $(x_n')_n$ of norm-1 functionals in E' and the convex function $f(x) = \sum_n n(|x_n'(x)| - 1/2)^+$. Then notice that f is finite and continuous if and only if $(x_n')_n$ is weak star convergent to 0.]

5. Suppose that f is a convex function defined on an open ball $B_r(a)$ in $\mathbb{R}^n$. Prove that there exists a constant $C > 0$, depending only on n, such that

$$\sup_{\overline{B}_{r/2}(a)} |f| \leq \frac{C}{m(B_r(a))} \int_{B_r(a)} |f(y)|\, dy$$

and

$$\operatorname*{ess\,sup}_{\overline{B}_{r/2}(a)} |Df| \leq \frac{C}{r \cdot m(B_r(a))} \int_{B_r(a)} |f(y)|\, dy.$$

6. (Two basic properties of infimal convolutions) Prove that:

(i) $(f \odot g)^* = f^* + g^*$;

(ii) $(f + g)^* = f^* \odot g^*$ if the effective domain of f contains a point of continuity of g.

7. Use the Moreau-Yosida approximates to complete the proof of the implication (i) $\Rightarrow$ (ii) in Theorem 3.7.8.
[*Hint*: In fact, $f^{**}(x) \geq \liminf_{\varepsilon \to 0} f_\varepsilon^{**}(x) = \liminf_{\varepsilon \to 0} f_\varepsilon(x) = f(x)$.]

3.9 Recognizing Convex Functions

We start with the following variant of Theorem 3.7.1:

Theorem 3.9.1 *Suppose that f is defined on an open convex set U in a Banach space. If f is convex on U and Gâteaux differentiable at $a \in U$, then*

$$f(x) \geq f(a) + f'(a\,; x - a) \quad \text{for every } x \in U. \tag{3.15}$$

If f is Gâteaux differentiable throughout U, then f is convex if and only if (3.15) *holds for all $a \in U$. Moreover, f is strictly convex if and only if the inequality is strict for $x \neq a$.*

On intervals, a differentiable function is convex if and only if its derivative is nondecreasing. The higher dimensional analogue of this fact is as follows:

Theorem 3.9.2 *Suppose that f is Gâteaux differentiable on the open convex set U in a Banach space. Then f is convex if and only if*

$$f'(x\,;x-y) \geq f'(y\,;x-y) \tag{3.16}$$

for all $x, y \in U$.

The variant of this result for strictly convex functions asks the above inequality to be strict for $x \neq y$ in U.

Proof. If f is convex, then for x and y in U and $0 < t < 1$ we have

$$\frac{f(y + t(x-y)) - f(y)}{t} \leq f(x) - f(y)$$

so by letting $t \to 0+$ we obtain $f'(y\,;x-y) \leq f(x) - f(y)$. Interchanging x and y, we also have $f'(x\,;y-x) \leq f(y) - f(x)$. Adding, we arrive at (3.16).

Suppose now that (3.16) holds. Let $x, y \in U$ and consider the function $g(\lambda) = f((1-\lambda)x + \lambda y)$, $\lambda \in [0,1]$. One can easily verify that

$$\lambda_1 \leq \lambda_2 \text{ implies } g'(\lambda_1) \leq g'(\lambda_2)$$

which shows that g is convex. Then

$$\begin{aligned} f((1-\lambda)x + \lambda y) &= g(\lambda) = g(\lambda \cdot 1 + (1-\lambda) \cdot 0) \\ &\leq \lambda g(1) + (1-\lambda) g(0) = (1-\lambda) f(x) + \lambda f(y). \end{aligned}$$

□

When the ambient space is $\mathbb{R}^n$, then the inequality (3.16) becomes

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0. \tag{3.17}$$

In this context, a function $F\colon U \to \mathbb{R}^n$ is said to be *nondecreasing* (respectively *increasing*) if it is the gradient of a convex (strictly convex) function.

Higher differentiability leads to other important criteria of convexity.

Suppose that $f\colon U \to \mathbb{R}$ is Gâteaux differentiable. We say that f is *twice Gâteaux differentiable* at $a \in U$ if the limit

$$f''(a\,;v,w) = \lim_{\lambda \to 0} \frac{f'(a + \lambda w, v) - f'(a\,;v)}{\lambda}$$

exists for all v, w in the ambient Banach space E. This gives rise to a map $f''(a)\colon (v,w) \mapsto f''(a\,;v,w)$, from $E \times E$ into $\mathbb{R}$, called the *second Gâteaux differential* of f at a. One can prove easily that this function is homogeneous in v and w, that is,

$$f''(a\,;\lambda v, \mu w) = \lambda \mu f''(a\,;v,w)$$

for all $\lambda, \mu \in \mathbb{R}$. Another immediate fact is as follows:

Lemma 3.9.3 *If $f\colon U \to \mathbb{R}$ is twice differentiable, then it is also twice Gâteaux differentiable and*

$$d^2 f(a)(v,w) = f''(a\,; v,w) \tag{3.18}$$

for all $a \in U$ and $v,w \in E$.

Our next goal is to establish the analogue of Taylor's formula in the context of Gâteaux differentiability and to infer from it an important characterization of convexity under the presence of Gâteaux differentiability.

Theorem 3.9.4 (Taylor's formula) *If f is twice Gâteaux differentiable at all points of the segment $[a, a+v]$ relative to the pair (v,v), then there exists a $\theta \in (0,1)$ such that*

$$f(a+v) = f(a) + f'(a\,; v) + \frac{1}{2} f''(a+\theta v\,; v,v). \tag{3.19}$$

Proof. Consider the function $g(t) = f(a+tv)$, for $t \in [0,1]$. Its derivative is

$$\begin{aligned} g'(t) &= \lim_{\varepsilon \to 0} \frac{g(t+\varepsilon) - g(t)}{\varepsilon} \\ &= \lim_{\varepsilon \to 0} \frac{f(a+tv+\varepsilon v) - f(a+tv)}{\varepsilon} = f'(a+tv\,; v) \end{aligned}$$

and similarly, $g''(t) = f''(a+tv\,; v,v)$. Then by the usual Taylor's formula we get a $\theta \in (0,1)$ such that

$$g(1) = g(0) + g'(0) + \frac{1}{2}\, g''(\theta),$$

which in turn yields the formula (3.19). $\square$

Corollary 3.9.5 *Suppose that f is twice Gâteaux differentiable on the open convex set U in a Banach space E and*

$$f''(a\,; v,v) \geq 0 \quad \textit{for all } a \in U,\ v \in E. \tag{3.20}$$

Then f is convex on U. If the above inequality is strict for $v \neq 0$, then f is strictly convex.

Proof. In fact, by Taylor's formula we have

$$f(x) = f(a) + f'(a\,; x-a) + \frac{1}{2}\, f''(a+\theta(x-a)\,; x-a, x-a)$$

for some $\theta \in (0,1)$, so by our hypothesis,

$$f(x) \geq f(a) + f'(a\,; x-a)$$

and the conclusion follows from Theorem 3.9.1. □

When $E = \mathbb{R}^n$ and $f''(a; v, w)$ is bilinear, it is easy to check the equality

$$f''(a\,; v, w) = \langle (\mathrm{Hess}_a\, f)v, w\rangle,$$

where

$$\mathrm{Hess}_a\, f = \Big(\frac{\partial^2 f}{\partial x_i \partial x_j}(a)\Big)_{i,j=1}^{n}$$

is the *Hessian matrix* of f at a.

Corollary 3.9.5 shows that the positivity (strict positivity) of the Hessian matrix at all points of U guarantees the convexity (strict convexity) of f.

If $A \in \mathrm{M}_n(\mathbb{R})$ is a strictly positive matrix and $u \in \mathbb{R}^n$, then the function

$$f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle x, u\rangle,$$

satisfies

$$f'(x\,; v) = \langle v, Ax\rangle - \langle v, u\rangle,$$
$$f''(x\,; v, w) = \langle Av, w\rangle = \langle v, Aw\rangle,$$

so by Corollary 3.9.5 it follows that f is strictly convex. By Theorem 3.4.5, f admits a global minimum a. According to Fermat's theorem (applied to the function $t \to f(a + tv)$), we infer that $f'(a\,; v) = 0$ for all v. This shows that a is the solution of the equation

$$Ax = u.$$

The above idea, to solve equations by finding the minimum of suitable functionals, is very useful in partial differential equations. See Appendix C.

Exercises

1. Consider the open set $A = \{(x, y, z) \in \mathbb{R}^3 \mid x, y > 0, xy > z^2\}$. Prove that A is convex and the function

$$f\colon A \to \mathbb{R}, \quad f(x, y, z) = \frac{1}{xy - z^2}$$

is strictly convex. Then, infer the inequality

$$\frac{8}{(x_1 + x_2)(y_1 + y_2) - (z_1 + z_2)^2} < \frac{1}{x_1y_1 - z_1^2} + \frac{1}{x_2y_2 - z_2^2}$$

which works for every pair of distinct points (x_1, y_1, z_1) and (x_2, y_2, z_2) of the set A.
[*Hint*: Compute the Hessian of f.]

2. (Minkowski's inequality for $p = 0$) Use calculus to prove that the function

$$f\colon [0,\infty)^n \to \mathbb{R}, \quad f(x_1,\dots,x_n) = \sqrt[n]{x_1\cdots x_n}$$

is concave and infer the inequality

$$\sqrt[n]{(x_1+y_1)\cdots(x_n+y_n)} \geq \sqrt[n]{x_1\cdots x_n} + \sqrt[n]{y_1\cdots y_n}$$

which works for all $x_1,\dots,x_n,y_1,\dots,y_n \geq 0$.
[*Hint*: Notice that $\sum_{i,j=1}^n \frac{\partial^2 F}{\partial x_i \partial x_j}(x_1,\dots,x_n)h_i h_j \geq 0$ for all $h_1,\dots,h_n$ in $\mathbb{R}$.]

3. Prove that the function $f(x,y) = y^2/(1-|x|)$ is convex and bounded on the open unit disc $D_1(0) = \{(x,y) \in \mathbb{R}^2 \mid x^2+y^2 < 1\}$.
[*Hint*: The function f can be represented as the maximum of two convex functions, $y^2/(1-x)$ and $y^2/(1+x)$.]

4. Suppose that f is a convex function on an open convex set in $\mathbb{R}^n$. If f is twice continuously differentiable, prove that its Hessian matrix is positive at each point of U.
[*Hint*: See the formula

$$\langle \nabla f(x) - \nabla f(y), x-y\rangle = \int_0^1 \langle H_f((1-t)x+ty)(x-y), x-y\rangle \, dt. \quad]$$

3.10 The Convex Programming Problem

The aim of this section is to discuss the problem of minimizing a convex function over a convex set defined by a system of convex inequalities. The main result is the equivalence of this problem to the so-called saddle-point problem. Assuming the differentiability of the functions concerned, the solution of the saddle-point problem is characterized by the Karush–Kuhn–Tucker conditions, which will be made explicit in Theorem 3.10.2 below.

In what follows $f, g_1, \dots, g_m$ will denote convex functions on $\mathbb{R}^n$. The *convex programming problem* for these data is to minimize $f(x)$ over the convex set

$$X = \{x \in \mathbb{R}^n \mid x \geq 0,\, g_1(x) \leq 0, \dots, g_m(x) \leq 0\}.$$

In optimization theory f represents a *cost*, which is minimized over the *feasible set* X.

A particular case is the standard *linear programming problem*. In this problem we seek to maximize a linear function

$$L(x) = -\langle x, c\rangle = -\sum_{k=1}^n c_k x_k$$

subject to the constraints

$$x \geq 0 \quad \text{and} \quad Ax \leq b.$$

Here $A \in \mathrm{M}_n(\mathbb{R})$ and $b, c \in \mathbb{R}^n$. Notice that this problem can be easily converted into a minimization problem, by replacing L by $-L$. According to Theorem 3.4.7, L attains its global maximum at an extreme point of the convex set $\{x \mid x \geq 0, Ax \leq b\}$. This point can be found by the *simplex algorithm* of G. B. Dantzig. See [212] for details.

Linear programming has many applications in industry and banking, which explains the great interest in faster algorithms. Let us mention here that in 1979, L. V. Khachian invented an algorithm having a polynomial computing time of order $\leq Kn^6$, where K is a constant.

Linear programming is also able to solve theoretical problems. The following example is due to E. Stiefel: Consider a matrix $A = (a_{ij})_{i,j} \in \mathrm{M}_{m \times n}(\mathbb{R})$ and a vector $b \in \mathbb{R}^m$ such that the system $Ax = b$ has no solution. Typically this occurs when we have more equations than unknowns. The error in the equation of rank i is a function of the form

$$e_i(x) = \sum_{j=1}^{n} a_{ij} x_j - b_i.$$

The problem of Chebyshev approximation is to minimize the maximum absolute error

$$X = \max\{|e_i(x)| \mid i = 1, \ldots, m\}.$$

Letting X be a new unknown, this problem can be read as

$$\text{minimize } X$$

subject to the inequalities

$$-X \leq \sum_{j=1}^{n} a_{ij} x_j - b_i \leq X \quad (i = 1, \ldots, m),$$

which can be easily converted into a standard linear programming problem.

We pass now to the convex programming problem. As in the case of any constrained extremal problem, one can apply the method of Lagrange multipliers in order to eliminate the constraints (at the cost of increasing the number of variables). The *Lagrangian function* associated with the convex programming problem is the function

$$F(x, y) = f(x) + y_1 g_1(x) + \cdots + y_m g_m(x)$$

of $n + m$ real variables $x_1, \ldots, x_n, y_1, \ldots, y_m$ (the components of x and respectively of y). A *saddle point* of F is any point (x^0, y^0) of $\mathbb{R}^n \times \mathbb{R}^m$ such that

$$x^0 \geq 0, \quad y^0 \geq 0$$

and

$$F(x^0, y) \leq F(x^0, y^0) \leq F(x, y^0)$$

for all $x \geq 0$, $y \geq 0$. The saddle points of F will provide solutions to the convex programming problem that generates F:

Theorem 3.10.1 *Let (x^0, y^0) be a saddle point of the Lagrangian function F. Then x^0 is a solution to the convex programming problem and*

$$f(x^0) = F(x^0, y^0).$$

Proof. The condition $F(x^0, y) \leq F(x^0, y^0)$ yields

$$y_1 g_1(x^0) + \cdots + y_m g_m(x^0) \leq y_1^0 g_1(x^0) + \cdots + y_m^0 g_m(x^0).$$

By keeping $y_2, \ldots, y_m$ fixed and taking the limit as $y_1 \to \infty$ we infer that $g_1(x^0) \leq 0$. Similarly, $g_2(x^0) \leq 0, \ldots, g_m(x^0) \leq 0$. Thus x^0 belongs to the feasible set X.

From $F(x^0, 0) \leq F(x^0, y^0)$ and the definition of X we infer

$$0 \leq y_1^0 g_1(x^0) + \cdots + y_m^0 g_m(x^0) \leq 0,$$

that is, $y_1^0 g_1(x^0) + \cdots + y_m^0 g_m(x^0) = 0$. Then $f(x^0) = F(x^0, y^0)$. Since $F(x^0, y^0) \leq F(x, y^0)$ for all $x \geq 0$, we have

$$f(x^0) \leq f(x) + y_1^0 g_1(x) + \cdots + y_m^0 g_m(x) \leq f(x)$$

for all $x \geq 0$, which shows that x^0 is a solution to the convex programming problem. □

Theorem 3.10.2 (The Karush–Kuhn–Tucker conditions) *Suppose that the convex functions $f, g_1, \ldots, g_m$ are differentiable on $\mathbb{R}^n$. Then (x^0, y^0) is a saddle point of the Lagrangian function F if and only if*

$$x^0 \geq 0, \tag{3.21}$$

$$\frac{\partial F}{\partial x_k}(x^0, y^0) \geq 0, \quad \textit{for } k = 1, \ldots, n, \tag{3.22}$$

$$\frac{\partial F}{\partial x_k}(x^0, y^0) = 0 \quad \textit{whenever } x_k^0 > 0, \tag{3.23}$$

and

$$y^0 \geq 0, \tag{3.24}$$

$$\frac{\partial F}{\partial y_j}(x^0, y^0) = g_j(x^0) \leq 0, \quad \textit{for } j = 1, \ldots, m, \tag{3.25}$$

$$\frac{\partial F}{\partial y_j}(x^0, y^0) = 0 \quad \textit{whenever } y_j^0 > 0. \tag{3.26}$$

Proof. If (x^0, y^0) is a saddle point of F, then (3.21) and (3.24) are clearly fulfilled. Also,

$$F(x^0 + te_k, y^0) \geq F(x^0, y^0) \quad \text{for all } t \geq -x_k^0.$$

If $x_k^0 = 0$, then

$$\frac{\partial F}{\partial x_k}(x^0, y^0) = \lim_{t \to 0+} \frac{F(x^0 + te_k, y^0) - F(x^0, y^0)}{t} \geq 0.$$

If $x_k^0 > 0$, then $\frac{\partial F}{\partial x_k}(x^0, y^0) = 0$ by Fermat's theorem. In a similar way one can prove (3.25) and (3.26).

Suppose now that the conditions (3.21)–(3.26) are satisfied. As $F(x, y^0)$ is a differentiable convex function of x (being a linear combination, with positive coefficients, of such functions), it verifies the assumptions of Theorem 3.9.1. Taking into account the conditions (3.21)–(3.23), we are led to

$$\begin{aligned} F(x, y^0) &\geq F(x^0, y^0) + \langle x - x^0, \nabla_x F(x^0, y^0) \rangle \\ &= F(x^0, y^0) + \sum_{k=1}^{n} (x_k - x_k^0) \frac{\partial F}{\partial x_k}(x^0, y^0) \\ &= F(x^0, y^0) + \sum_{k=1}^{n} x_k \frac{\partial F}{\partial x_k}(x^0, y^0) \geq F(x^0, y^0) \end{aligned}$$

for all $x \geq 0$. On the other hand, by (3.25)–(3.26), for $y \geq 0$, we have

$$\begin{aligned} F(x^0, y) &= F(x^0, y^0) + \sum_{j=1}^{m} (y_j - y_j^0)\, g_j(x^0) \\ &= F(x^0, y^0) + \sum_{j=1}^{m} y_j g_j(x^0) \\ &\leq F(x^0, y^0). \end{aligned}$$

Consequently, (x^0, y^0) is a saddle point of F. $\square$

We shall illustrate Theorem 3.10.2 by the following example:

$$\text{minimize } (x_1 - 2)^2 + (x_2 + 1)^2 \quad \text{subject to } 0 \leq x_1 \leq 1 \text{ and } 0 \leq x_2 \leq 2.$$

Here $f(x_1, x_2) = (x_1 - 2)^2 + (x_2 + 1)^2$, $g_1(x_1, x_2) = x_1 - 1$ and $g_2(x_1, x_2) = x_2 - 2$. The Lagrangian function attached to this problem is

$$F(x_1, x_2, y_1, y_2) = (x_1 - 2)^2 + (x_2 + 1)^2 + y_1(x_1 - 1) + y_2(x_2 - 2)$$

and the Karush–Kuhn–Tucker conditions give us the equations

$$\begin{cases} x_1(2x_1 - 4 + y_1) = 0, \\ x_2(2x_2 + 2 + y_2) = 0, \\ y_1(x_1 - 1) = 0, \\ y_2(x_2 - 2) = 0, \end{cases} \tag{3.27}$$

and the inequalities

$$\begin{cases} 2x_1 - 4 + y_1 \geq 0, \\ 2x_2 + 2 + y_2 \geq 0, \\ 0 \leq x_1 \leq 1 \text{ and } 0 \leq x_2 \leq 2, \\ y_1, y_2 \geq 0. \end{cases} \tag{3.28}$$

The system of equations (3.27) admits 9 solutions: $(1, 0, 2, 0)$, $(1, 2, 2, -6)$, $(1, -1, 2, 0)$, $(0, 0, 0, 0)$, $(2, 0, 0, 0)$, $(0, -1, 0, 0)$, $(2, -1, 0, 0)$, $(0, 0, 0, -1)$ and $(2, 0, 0, -1)$, of which only $(1, 0, 2, 0)$ verifies also the inequalities (3.28). Consequently,

$$\inf_{\substack{0 \leq x_1 \leq 1 \\ 0 \leq x_2 \leq 2}} f(x_1, x_2) = f(1, 0) = 2.$$

We next indicate a fairly general situation when the convex programming problem is equivalent to the saddle-point problem. For this we shall need the following technical result, known as Farkas' lemma:

Lemma 3.10.3 *Let $f_1, \ldots, f_m$ be convex functions defined on a nonempty convex set Y in $\mathbb{R}^n$. Then either there exists y in Y such that $f_1(y) < 0, \ldots, f_m(y) < 0$, or there exist nonnegative numbers $a_1, \ldots, a_m$, not all zero, such that*

$$a_1 f_1(y) + \cdots + a_m f_m(y) \geq 0 \quad \text{for all } y \in Y.$$

Proof. Assume that the first alternative does not work and consider the set

$$C = \Big\{ (t_1, \ldots, t_m) \in \mathbb{R}^m \mid \text{there is } y \in Y \text{ with } f_k(y) < t_k \\ \text{for all } k = 1, \ldots, m \Big\}.$$

Then C is an open convex set that does not contain the origin of $\mathbb{R}^m$. According to Theorem 3.3.1, C and the origin can be separated by a closed hyperplane, that is, there exist scalars $a_1, \ldots, a_m$ not all zero, such that for all $y \in Y$ and all $\varepsilon_1, \ldots, \varepsilon_m > 0$,

$$a_1(f_1(y) + \varepsilon_1) + \cdots + a_m(f_m(y) + \varepsilon_m) \geq 0. \tag{3.29}$$

Keeping $\varepsilon_2, \ldots, \varepsilon_m$ fixed and letting $\varepsilon_1 \to \infty$, we infer that $a_1 \geq 0$. Similarly, $a_2 \geq 0, \ldots, a_m \geq 0$. Letting $\varepsilon_1 \to 0, \ldots, \varepsilon_m \to 0$ in (3.29) we conclude that $a_1 f_1(y) + \cdots + a_m f_m(y) \geq 0$ for all y in Y. $\square$

Theorem 3.10.4 (Slater's condition) *Suppose that x^0 is a solution of the convex programming problem. If there exists $x^* \geq 0$ such that $g_1(x^*) < 0, \dots, g_m(x^*) < 0$, then one can find a y^0 in $\mathbb{R}^m$ for which (x^0, y^0) is a saddle point of the associated Lagrangian function F.*

Proof. By Lemma 3.10.3, applied to the functions $g_1, \dots, g_m, f - f(x^0)$ and the set $Y = \mathbb{R}^n_+$, we can find $a_1, \dots, a_m, a_0 \geq 0$, not all zero, such that

$$a_1 g_1(x) + \cdots + a_m g_m(x) + a_0(f(x) - f(x^0)) \geq 0 \tag{3.30}$$

for all $x \geq 0$. A moment's reflection shows that $a_0 > 0$. Put $y_j^0 = a_j/a_0$ and $y^0 = (y_1^0, \dots, y_m^0)$. By (3.30) we infer that $f(x^0) \leq f(x) + \sum_{j=1}^m y_j^0 g_j(x) = F(x, y^0)$ for all $x \geq 0$. Particularly, for $x = x^0$, this yields

$$f(x^0) \leq f(x^0) + \sum_{j=1}^m y_j^0 g_j(x^0) \leq f(x^0)$$

that is, $\sum_{j=1}^m y_j^0 g_j(x^0) = 0$, whence $F(x^0, y^0) = f(x^0) \leq F(x, y^0)$ for all $x \geq 0$. On the other hand, for $y \geq 0$ we have

$$F(x^0, y^0) = f(x^0) \geq f(x^0) + \sum_{j=1}^m y_j g_j(x^0) = F(x^0, y),$$

so that (x^0, y^0) is a saddle point. $\square$

We end this section with a nice geometric application of convex programming (more precisely, of quadratic programming), which was noted by J. Franklin, in his beautiful introduction to mathematical methods of economics [82]. It is about a problem of J. Sylvester, requiring the least circle which contains a given set of points in the plane.

Suppose the given points are $a_1, \dots, a_m$. They lie inside the circle of center x and radius r if

$$\|a_k - x\|^2 \leq r^2 \quad \text{for } k = 1, \dots, m. \tag{3.31}$$

We want to find x and r so as to minimize r. Letting

$$x_0 = \frac{1}{2}(r^2 - \|x\|^2),$$

we can replace the quadratic constraints (3.31) by linear ones,

$$x_0 + \langle a_k, x \rangle \geq b_k \quad \text{for } k = 1, \dots, m.$$

Here $b_k = \|a_k\|^2/2$. In this way, Sylvester's problem becomes a problem of quadratic programming,

$$\text{minimize } 2x_0 + x_1^2 + x_2^2,$$

subject to the m linear inequalities

$$x_0 + a_{k1}x_1 + a_{k2}x_2 \geq b_k \quad (k = 1, \dots, m).$$

A numerical algorithm to solve problems of this kind is described in [51].

Exercises

1. Minimize $x^2 + y^2 - 6x - 4y$, subject to $x \geq 0$, $y \geq 0$ and $x^2 + y^2 \leq 1$.

2. Infer from Farkas' lemma the *fundamental theorem of Markov processes*: Suppose that $(p_{ij})_{i,j} \in \mathrm{M}_n(\mathbb{R})$ is a matrix with nonnegative coefficients and

$$\sum_{i=1}^{n} p_{ij} = 1 \quad \text{for all } j = 1, \ldots, n.$$

Then there exists a vector $x \in \mathbb{R}^n_+$ such that

$$\sum_{j=1}^{n} x_j = 1 \quad \text{and} \quad \sum_{j=1}^{n} p_{ij} x_j = x_i \quad \text{for all } i = 1, \ldots, n.$$

3. (A variant of Farkas' lemma) The following result is an analogue of the *Fredholm alternative* for linear inequalities. Let A be an $m \times n$ real matrix, and let b be a vector of $\mathbb{R}^m$. Prove that one, and only one, of the following two alternatives is true:

(i) the system $Ax = b$ has a solution $x \in \mathbb{R}^n_+$;
(ii) there exists a vector $y \in \mathbb{R}^n$ such that $A^\star y \in \mathbb{R}^m_+$ and $\langle y, b \rangle < 0$.

4. Suppose that C is a convex subset of $\mathbb{R}^n$ and $a \in C$. The *tangent cone* to C at a is the closed cone $T_C(a) = \overline{\mathbb{R}_+(C - a)}$, and the *normal cone* to C at a is the closed cone $N_C(a) = \{v \in \mathbb{R}^n \mid \langle v, x - a \rangle \leq 0 \text{ for all } x \in C\}$.

(i) Compute $T_C(a)$ and $N_C(a)$ when C is the unit disc.
(ii) Prove that the polar set of $T_C(a)$ is $N_C(a)$ (and vice versa).

5. Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a convex function and $C = \{x \mid f(x) \leq 0\}$. Assume there exists a point x such that $f(x) < 0$. Prove that

$$T_C(a) = \{v \mid f'(a; v) \leq 0\} \quad \text{and} \quad N_C(a) = \mathbb{R}_+ \partial f(a)$$

for all $a \in \mathbb{R}^n$ such that $f(a) = 0$.

6. (Self-dual cones) Suppose that C is one of the following cones: $\mathbb{R}^n_+$, $\mathrm{Sym}^+(n, \mathbb{R})$ and $\{x \in \mathbb{R}^n_+ \mid x_1^2 \geq x_2^2 + \cdots + x_n^2\}$. Prove that $N_C(0) = -C$.

7. Suppose that C is a convex subset of $\mathbb{R}^n$ and that $f\colon \mathbb{R}^n \to \mathbb{R}$ is a convex function. Prove that the following assertions are equivalent for $a \in C$:

(i) a is a minimizer for $f|_C$;
(ii) $f'(x; v) \geq 0$ for all $v \in T_C(a)$;
(iii) $0 \in \partial f(a) + N_C(a)$.

8. (The Karush–Kuhn–Tucker conditions) Suppose that $f, g_1, \ldots, g_m$ are convex functions on $\mathbb{R}^n$ and that there is a point $x \in \mathbb{R}^n$ such that $g_i(x) < 0$ for $i = 1, \ldots, m$. Then $a \in \mathbb{R}^n$ is a solution of the convex programming problem for these data if and only if there is a vector $w = (w_1, \ldots, w_m) \in \mathbb{R}^m_+$ such that

$$0 \in \partial f(a) + w_1 \partial g_1(a) + \cdots + w_m \partial g_m(a)$$
$$g_i(a) \le 0, \quad w_i g_i(a) = 0 \quad \text{for } i = 1, \ldots, m.$$

3.11 Fine Properties of Differentiability

The aim of this section is to prove two remarkable results which play a considerable role in convex analysis: Rademacher's theorem (asserting the differentiability of Lipschitz functions almost everywhere) and Alexandrov's theorem (asserting the twice differentiability of convex functions almost everywhere).

Theorem 3.11.1 (Rademacher's theorem) *Every locally Lipschitz function $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is almost everywhere differentiable.*

Proof. Since a vector-valued function f is differentiable at a point if and only if all of its components are differentiable at that point, we may restrict ourselves to the case of real-valued functions. Also, since differentiability is a local property, we may as well assume that f is Lipschitz. See Kirszbraun's Theorem 3.7.4.

The remainder of the proof will be done in three steps.

Step 1. Fix arbitrarily a vector $v \in \mathbb{R}^n$, $\|v\| = 1$. We shall show that the directional derivative $f'(x\,;v)$ exists for almost every $x \in \mathbb{R}^n$. In fact, for each $x \in \mathbb{R}^n$ consider the limits

$$\underline{D}f(x\,;v) = \liminf_{t\to 0} \frac{f(x+tv) - f(x)}{t}$$

and

$$\overline{D}f(x\,;v) = \limsup_{t\to 0} \frac{f(x+tv) - f(x)}{t},$$

which exist in $\overline{\mathbb{R}}$. The set

$$E_v = \{x \in \mathbb{R}^n \mid \underline{D}f(x\,;v) < \overline{D}f(x\,;v)\}$$

equals the set where the directional derivative $f'(x\,;v)$ does not exist. As in the proof of Theorem 3.8.3 we may conclude that E_v is Lebesgue measurable. We shall show that E_v is actually Lebesgue negligible. In fact, by Lebesgue's theory on the differentiability of absolutely continuous functions (see [74] or [103]) we infer that the functions

$$g(t) = f(x+tv), \quad t \in \mathbb{R}$$

are differentiable almost everywhere. This implies that the Lebesgue measure of the intersection of E_v with any line L is Lebesgue negligible. Then, by Fubini's theorem, we conclude that E_v is itself Lebesgue negligible.

Step 2. According to the discussion above we know that

$$\nabla f(x) = \Big(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x)\Big)$$

exists almost everywhere. We shall show that

$$f'(x; v) = \langle v, \nabla f(x)\rangle$$

for almost every $x \in \mathbb{R}^n$. In fact, for an arbitrary fixed $\varphi \in C_c^\infty(\mathbb{R}^n)$ we have

$$\int_{\mathbb{R}^n} \Big[\frac{f(x+tv) - f(x)}{t}\Big]\varphi(x)\,dx = -\int_{\mathbb{R}^n} f(x)\Big[\frac{\varphi(x) - \varphi(x-tv)}{t}\Big]\,dx.$$

Since

$$\Big|\frac{f(x+tv) - f(x)}{t}\Big| \leq \mathrm{Lip}(f),$$

we can apply the dominated convergence theorem to get

$$\int_{\mathbb{R}^n} f'(x; v)\varphi(x)\,dx = -\int_{\mathbb{R}^n} f(x)\varphi'(x; v)\,dx.$$

By taking into account Fubini's theorem and the absolute continuity of f on lines we can continue as follows:

$$\begin{aligned}
-\int_{\mathbb{R}^n} f(x)\varphi'(x; v)\,dx &= -\sum_{k=1}^n v_k \int_{\mathbb{R}^n} f(x)\frac{\partial \varphi}{\partial x_k}(x)\,dx \\
&= \sum_{k=1}^n v_k \int_{\mathbb{R}^n} \frac{\partial f}{\partial x_k}(x)\varphi(x)\,dx \\
&= \int_{\mathbb{R}^n} \langle v, \nabla f(x)\rangle\varphi(x)\,dx
\end{aligned}$$

and this leads us to the formula $f'(x; v) = \langle v, \nabla f(x)\rangle$, as φ was arbitrarily fixed.

Step 3. Consider now a countable family $(v_i)_i$ of unit vectors, which is dense in the unit sphere of $\mathbb{R}^n$. By the above reasoning we infer that the complement of each of the sets

$$A_i = \{x \in \mathbb{R}^n \mid Df(x; v_i) \text{ and } \nabla f(x) \text{ exist and } f'(x; v_i) = \langle v_i, \nabla f(x)\rangle\}$$

is Lebesgue negligible, and thus the same is true for the complement of

$$A = \bigcap_{i=1}^\infty A_i.$$

We shall show that f is differentiable at all points of A. This will be done by considering the function

$$R(x, v, t) = \frac{f(x+tv) - f(x)}{t} - \langle v, \nabla f(x)\rangle,$$

for $x \in A$, $v \in \mathbb{R}^n$, $\|v\| = 1$, and $t \in \mathbb{R}\backslash\{0\}$. Since

$$\begin{aligned} |R(x,v,t) - R(x,v',t)| &\leq \operatorname{Lip}(f) \cdot \|v - v'\| + \|\nabla f(x)\| \cdot \|v - v'\| \\ &\leq (\sqrt{n}+1) \operatorname{Lip}(f) \cdot \|v - v'\|, \end{aligned}$$

the function $R(x,v,t)$ is Lipschitz in v.

Suppose there are given a point $a \in A$ and a number $\varepsilon > 0$. Since the unit sphere of $\mathbb{R}^n$ is compact and the family $(v_i)_i$ is a dense subset, we can choose a natural number N such that

$$\inf_{i \in \{0,\ldots,N\}} \|v - v_i\| < \frac{\varepsilon}{2(\sqrt{n}+1)\operatorname{Lip}(f)}$$

for all v with $\|v\| = 1$. By the definition of A, there exists a $\delta > 0$ such that

$$|R(a, v_i, t)| < \frac{\varepsilon}{2}$$

for all $i \in \{0, \ldots, N\}$ and all $|t| < \delta$. Then

$$\begin{aligned} |R(a,v,t)| &\leq \inf_{i \in \{0,\ldots,N\}} (|R(a,v_i,t)| + |R(a,v_i,t) - R(a,v,t)|) \\ &\leq \frac{\varepsilon}{2} + (\sqrt{n}+1)\operatorname{Lip}(f) \cdot \frac{\varepsilon}{2(\sqrt{n}+1)\operatorname{Lip}(f)} \\ &= \varepsilon \end{aligned}$$

for all v in the unit sphere of $\mathbb{R}^n$ and all t with $|t| < \delta$. This assures the differentiability of f at a. □

Rademacher's theorem allows us to extend a number of important results such as the area formula, the change of variable formula, and the invariance of Sobolev spaces under Lipschitz isomorphisms, from the context of smooth functions to that of Lipschitz functions. See L. C. Evans and R. F. Gariepy [74], and W. P. Ziemer [252].

We pass now to the problem of almost everywhere second differentiability of convex functions. In order to simplify exposition we shall make constant use of *Landau's symbol o*, where

$$f = o(g) \text{ for } x \to 0 \text{ means } f = hg \text{ with } \lim_{x \to 0} h(x) = 0.$$

Theorem 3.11.2 (A. D. Alexandrov [5]) *Every convex function f on $\mathbb{R}^n$ is twice differentiable almost everywhere in the following sense: f is twice differentiable at a, with Alexandrov Hessian $\nabla^2 f(a)$ in $\operatorname{Sym}^+(n,\mathbb{R})$, if $\nabla f(a)$ exists, and if for every $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$\|x - a\| < \delta \textit{ implies } \sup_{y \in \partial f(x)} \|y - \nabla f(a) - \nabla^2 f(a)(x-a)\| \leq \varepsilon \|x - a\|.$$

Moreover, if a is such a point, then

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - \langle \nabla f(a), h \rangle - \frac{1}{2}\langle \nabla^2 f(a)h, h \rangle}{\|h\|^2} = 0.$$

Proof. By Theorem 3.7.1, the domain of the subdifferential ∂f is the whole space $\mathbb{R}^n$, while Theorem 3.8.3 shows that $\partial f(x) = \{df(x)\}$ for all x in

$$X_1 = \{x \in \mathbb{R}^n \mid f \text{ is differentiable at } x\},$$

which is a set whose complement is a negligible set.

We shall show that for almost all x in $\mathbb{R}^n$ there exists a matrix A in $\mathrm{Sym}^+(n, \mathbb{R})$ such that

$$df(y) = df(x) + A(y - x) + o(\|y - x\|) \quad \text{for all } y \in X_1. \tag{3.32}$$

We need the fact that $J = (I + \partial f)^{-1}$ is a nonexpansive map of $\mathbb{R}^n$ into itself. See Corollary 3.7.6. This yields a new set,

$$X_2 = \{J(x) \mid J \text{ is differentiable at } x \text{ and } dJ(x) \text{ is nonsingular}\},$$

whose complement is also a negligible set. In fact, by Rademacher's theorem, J is differentiable almost everywhere. Since J is Lipschitz, we may apply the area formula (see L. C. Evans and R. F. Gariepy [74, Theorem 3.3.2, p. 96]) to get

$$\int_B |\det(dJ(x))|\, dx = \int_{\mathbb{R}^n} \#(B \cap J^{-1}(y))\, dy \quad \text{for all Borel sets } B \text{ in } \mathbb{R}^n,$$

where $\#$ is the counting measure. By this formula (and the fact that J is onto) we infer that the complementary set of

$$\{x \mid J \text{ is differentiable at } x \text{ and } dJ(x) \text{ is nonsingular}\}$$

is a negligible set. On the other hand, any Lipschitz function maps negligible sets into negligible sets. See [218, Lemma 7.25]. Hence X_2 is indeed a set whose complementary set is negligible.

We shall show that the formula (3.32) works for all x in $X_3 = X_1 \cap X_2$ (which is a set with negligible complementary set). Our argument is based on the following fact concerning the solvability of nonlinear equations in $\mathbb{R}^n$: If $F\colon \overline{B}_\delta(0) \to \mathbb{R}^n$ is continuous, $0 < \varepsilon < \delta$ and $\|F(x) - x\| < \varepsilon$ for all $x \in \mathbb{R}^n$ with $\|x\| = \delta$, then $F(B_\delta(0)) \supset B_{\delta-\varepsilon}(0)$. See W. Rudin [218, Lemma 7.23] for a proof based on the Brouwer fixed point theorem.

By the definition of J,

$$df(J(x)) = x - J(x)$$

for all x with $J(x) \in X_3$ ($\subset X_1$). Suppose that $J(x) + \tilde{y} \in X_1$, where $\tilde{y}$ is small. Since J is Lipschitz and $dJ(x)$ is nonsingular, if $\tilde{y}$ is sufficiently small, then there exists an $\tilde{x}$ such that $J(x + \tilde{x}) = J(x) + \tilde{y}$. Moreover, we may choose $\tilde{x}$ to verify $\|\tilde{x}\| \le C\|\tilde{y}\|$ for some constant $C > 0$. Use a remark above (on the solvability of nonlinear equations) and the fact that $J(x+h) =$

$J(x) + dJ(x)h + o(\|h\|)$. Since J is nonexpansive, we also have $\|\tilde{y}\| \leq \|\tilde{x}\|$, hence $\|\tilde{x}\|$ and $\|\tilde{y}\|$ are comparable. Then

$$\begin{aligned} df(J(x) + \tilde{y}) &= df(J(x + \tilde{x})) = x + \tilde{x} - J(x + \tilde{x}) \\ &= df(J(x)) + (I - dJ(x))\tilde{x} + o(\|\tilde{x}\|). \end{aligned}$$

Due to the relation $J(x)+\tilde{y} = J(x)+dJ(x)\tilde{x}+o(\|\tilde{x}\|)$ and the comparability of $\|\tilde{x}\|$ and $\|\tilde{y}\|$, we have

$$\tilde{x} = (dJ(x))^{-1}\tilde{y} + o(\|\tilde{y}\|).$$

Hence $d(df)(J(x))$ exists and equals $(dJ(x))^{-1} - I$.

It remains to prove that

$$f(J(x) + \tilde{y}) = f(J(x)) + df(J(x))\tilde{y} + \frac{1}{2}\big\langle\big((dJ(x))^{-1} - I\big)\tilde{y}, \tilde{y}\big\rangle + o(\|\tilde{y}\|^2)$$

for $J(x) \in X_3$. Letting

$$R(\tilde{y}) = f(J(x) + \tilde{y}) - f(J(x)) - df(J(x))\tilde{y} - \frac{1}{2}\big\langle\big((dJ(x))^{-1} - I\big)\tilde{y}, \tilde{y}\big\rangle,$$

we get a locally Lipschitz function R such that $R(0) = 0$, and for almost all small $\tilde{y}$,

$$dR(\tilde{y}) = o(\|\tilde{y}\|).$$

By the mean value theorem we conclude that $R(\tilde{y}) = o(\|\tilde{y}\|^2)$ and the proof is complete. $\square$

The result of Theorem 3.11.2 can be easily extended to conclude that every proper convex function $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is twice differentiable almost everywhere on its effective domain.

Alexandrov's theorem has important applications to convex geometric analysis and partial differential equations. See the Comments at the end of this chapter.

Both Theorems 3.11.1 and 3.11.2 remain valid in the more general framework of semiconvex functions. A function f defined on a convex set in $\mathbb{R}^n$ is said to be *semiconvex* if $f + \lambda\|\cdot\|^2$ is a convex function for some $\lambda > 0$. An important example of a semiconvex function which is not necessarily convex is the Asplund function. See Exercise 2.

Exercises

1. (The existence of distributional derivatives) Suppose that $f\colon \mathbb{R}^n \to \mathbb{R}$ is a convex function. Prove that for all $i, j \in \{1, \ldots, n\}$ there exist signed Radon measures μ_{ij} (with $\mu_{ij} = \mu_{ji}$) such that

$$\int_{\mathbb{R}^n} f \frac{\partial^2 \varphi}{\partial x_i \partial x_j} \, dx = \int_{\mathbb{R}^n} \varphi \, d\mu_{ij} \quad \text{for every } \varphi \in C_c^2(\mathbb{R}^n).$$

Moreover, the measures μ_{ii} are nonnegative.
[*Hint*: Suppose first that f is smooth. For an arbitrarily fixed $\xi = (\xi_1, \dots, \xi_n) \in \mathbb{R}^n$, $\|\xi\| = 1$, consider the linear functional

$$L(\varphi) = \sum_{i,j=1}^n \int_{\mathbb{R}^n} f \frac{\partial^2 \varphi}{\partial x_i \partial x_j} \xi_i \xi_j \, dx, \quad \varphi \in C_c^\infty(\mathbb{R}^n).$$

Then $L(\varphi) \geq 0$ if $\varphi \geq 0$; this is clear if f is smooth, and in the general case we may use mollification. By a variant of the Riesz-Kakutani representation theorem (see [74], Corollary 1, pp. 53-54) we get a positive Radon measure μ^ξ such that $L(\varphi) = \int_{\mathbb{R}^n} \varphi \, d\mu^\xi$ for all $\varphi \in C_c^\infty(\mathbb{R}^n)$. Letting $e_1, \dots, e_n$ be the canonical basis of $\mathbb{R}^n$, we may choose $\mu_{ii} = \mu^{e_i}$ and

$$\mu_{ij} = \mu^{(e_i + e_j)/2} - 1/2 \cdot \mu^{e_i} - 1/2 \cdot \mu^{e_j} \quad \text{for } i \neq j. \quad]$$

2. (The Asplund function) Given a nonempty closed subset S of $\mathbb{R}^n$ we can associate to it the function

$$\varphi_S(x) = \frac{1}{2} \left(\|x\|^2 - d_S^2(x) \right),$$

where $d_S(x) = \inf\{\|x - s\| \mid s \in S\}$ represents the distance from x to S.

(i) Notice that $\varphi_S(x) = \sup\{\langle x, s\rangle - \frac{1}{2}\|s\|^2 \mid s \in S\}$ and infer that φ_S is a convex function.

(ii) Prove that φ_S is the conjugate of the function $f_S(x) = \|x\|^2/2$ if $x \in S$ and $f_S(x) = \infty$ otherwise.

(iii) Use Remark 3.7.9 (ii) to infer Bunt's theorem (that is, Theorem 3.2.2).

3. Let $f \colon \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz function and set

$$Z = \{x \mid f(x) = 0\}.$$

Prove that $df(x) = 0$ for almost every $x \in Z$.

4. Let $f, g \colon \mathbb{R}^n \to \mathbb{R}^m$ be locally Lipschitz functions and set

$$X = \{x \mid g(f(x)) = x\}.$$

Prove that $dg(f(x))df(x) = I$ for almost every $x \in X$.

5. (D. Cordero-Erausquin [59]) Suppose that f is a proper convex function with $\Omega = \operatorname{int}(\operatorname{dom}(f))$. We denote by $\Delta_A f$ the trace of the Alexandrov Hessian and by Δf the Laplacian of f in the sense of distributions. Prove that

$$\int_\Omega \varphi \Delta_A f \, dx \leq -\int_\Omega \nabla \varphi \cdot \nabla f \, dx$$

for all functions $\varphi \in C_c^\infty(\Omega)$, $f \geq 0$.

3.12 Prékopa–Leindler Type Inequalities

The aim of this section is to present several inequalities that play an important role in the geometric theory of convexity. The first one is related to the isoperimetric problem and represents a reverse form of the Rogers–Hölder inequality.

Theorem 3.12.1 (The Prékopa–Leindler inequality) *Let* $0 < \lambda < 1$ *and let* f, g, *and* h *be nonnegative integrable functions on* $\mathbb{R}^n$ *satisfying*

$$h((1-\lambda)x+\lambda y) \geq f(x)^{1-\lambda} g(y)^{\lambda}$$

for all $x, y \in \mathbb{R}^n$. *Then*

$$\int_{\mathbb{R}^n} h(x)\,dx \geq \left(\int_{\mathbb{R}^n} f(x)\,dx\right)^{1-\lambda} \left(\int_{\mathbb{R}^n} g(x)\,dx\right)^{\lambda}.$$

The connection with the Rogers–Hölder inequality will become clear after restating the above result in the form

$$\overline{\int}_{\mathbb{R}^n} \sup_{(1-\lambda)x+\lambda y=z} f(x)^{1-\lambda} g(y)^{\lambda}\,dz \geq \left(\int_{\mathbb{R}^n} f(x)\,dx\right)^{1-\lambda} \left(\int_{\mathbb{R}^n} g(x)\,dx\right)^{\lambda},$$

as h can be replaced by the supremum inside the left integral, and then passing to the more familiar form

$$\overline{\int}_{\mathbb{R}^n} \sup_{(1-\lambda)x+\lambda y=z} f(x)g(y)\,dz \geq \left(\int_{\mathbb{R}^n} f^p(x)\,dx\right)^{1/p} \left(\int_{\mathbb{R}^n} g^q(x)\,dx\right)^{1/q},$$

after replacing $1-\lambda$ by $1/p$, λ by $1/q$, f by f^p and g by g^q. The upper integral is used in the left-hand side because the integrand is not necessarily measurable.

As we shall show later, the Prékopa–Leindler inequality is just the particular case of a very general result, the Borell–Brascamp–Lieb inequality.

An important consequence of Theorem 3.12.1 is as follows:

Theorem 3.12.2 (Lusternik's general Brunn–Minkowski inequality) *Let* $s, t > 0$ *and let* X *and* Y *be nonempty bounded measurable sets in* $\mathbb{R}^n$ *such that* $sX + tY$ *is also measurable. Then*

$$\mathrm{Vol}_n(sX+tY)^{1/n} \geq s\,\mathrm{Vol}_n(X)^{1/n} + t\,\mathrm{Vol}_n(Y)^{1/n}.$$

Here Vol_n denotes the n-dimensional Lebesgue measure.

Proof. Since the Lebesgue measure Vol_n is positively homogeneous of degree n (that is, $\mathrm{Vol}_n(\alpha A) = \alpha^n \mathrm{Vol}_n(A)$ for every Borel set A and every $\alpha \geq 0$), we may restrict to the case where $s = 1-\lambda$ and $t = \lambda$ for some $\lambda \in (0,1)$.

Then we apply the Prékopa–Leindler inequality for $f = \chi_X$, $g = \chi_Y$ and $h = \chi_{(1-\lambda)X+\lambda Y}$, which yields

$$\begin{aligned}
\mathrm{Vol}_n((1-\lambda)X+\lambda Y) &= \int_{\mathbb{R}^n} \chi_{(1-\lambda)X+\lambda Y}(x)\,dx \\
&\geq \left(\int_{\mathbb{R}^n} \chi_X(x)\,dx\right)^{1-\lambda} \left(\int_{\mathbb{R}^n} \chi_Y(x)\,dx\right)^{\lambda} \\
&= \mathrm{Vol}_n(X)^{1-\lambda}\,\mathrm{Vol}_n(Y)^{\lambda}.
\end{aligned}$$

Applying this inequality for X replaced by $\mathrm{Vol}_n(X)^{-1/n}X$, Y replaced by $\mathrm{Vol}_n(Y)^{-1/n}Y$, and λ replaced by

$$\frac{\lambda\,\mathrm{Vol}_n(Y)^{1/n}}{(1-\lambda)\,\mathrm{Vol}_n(X)^{1/n}+\lambda\,\mathrm{Vol}_n(Y)^{1/n}},$$

we obtain

$$\mathrm{Vol}_n((1-\lambda)X+\lambda Y)^{1/n} \geq (1-\lambda)\,\mathrm{Vol}_n(X)^{1/n}+\lambda\,\mathrm{Vol}_n(Y)^{1/n},$$

which ends the proof. □

The hypothesis on the measurability of $sX + tY$ cannot be deduced from the measurability of X and Y. A counterexample can be found in a paper by W. Sierpiński [225].

The Brunn–Minkowski inequality represents the particular case of Theorem 3.12.2 above for convex bodies. A *convex body* is understood as a compact convex set in $\mathbb{R}^n$, with nonempty interior. In this case the measurability of the sets $sX + tY$ is automatic.

Theorem 3.12.3 (The Brunn–Minkowski inequality) *Let $\lambda \in (0,1)$ and let K and L be two convex bodies. Then*

$$\mathrm{Vol}_n((1-\lambda)K+\lambda L)^{1/n} \geq (1-\lambda)\,\mathrm{Vol}_n(K)^{1/n}+\lambda\,\mathrm{Vol}_n(L)^{1/n}.$$

Equality holds precisely when K and L are equal up to translation and dilation.

Theorem 3.12.3 says that the function $t \to \mathrm{Vol}_n((1-t)K+tL)^{1/n}$ is concave on $[0,1]$. It is also log-concave as follows from the AM–GM inequality.

The volume V of a ball $B_r(0)$ in $\mathbb{R}^3$ and the area S of its surface are connected by the relation

$$S = \frac{dV}{dR}.$$

This fact led H. Minkowski to define the *surface area* of a convex body K in $\mathbb{R}^n$ by the formula

$$S_{n-1}(K) = \lim_{\varepsilon\to 0+} \frac{\mathrm{Vol}_n(K+\varepsilon B)-\mathrm{Vol}_n(K)}{\varepsilon},$$

where B denotes the closed unit ball of $\mathbb{R}^n$. The agreement of this definition with the usual definition of the surface of a smooth surface is discussed in books by H. Federer [77] and Y. D. Burago and V. A. Zalgaller [45].

Theorem 3.12.4 (The isoperimetric inequality for convex bodies in $\mathbb{R}^n$) *Let K be a convex body in $\mathbb{R}^n$ and let B denote the closed unit ball of this space. Then*

$$\left(\frac{\mathrm{Vol}_n(K)}{\mathrm{Vol}_n(B)}\right)^{1/n} \leq \left(\frac{S_{n-1}(K)}{S_{n-1}(B)}\right)^{1/(n-1)}$$

with equality if and only if K is a ball.

Proof. In fact, by the Brunn–Minkowski inequality,

$$\begin{aligned}
S_{n-1}(K) &= \lim_{\varepsilon\to 0+} \frac{\mathrm{Vol}_n(K+\varepsilon B) - \mathrm{Vol}_n(K)}{\varepsilon} \\
&\geq \lim_{\varepsilon\to 0+} \frac{(\mathrm{Vol}_n(K)^{1/n} + \varepsilon \,\mathrm{Vol}_n(B)^{1/n})^n - \mathrm{Vol}_n(K)}{\varepsilon} \\
&= n \,\mathrm{Vol}_n(K)^{(n-1)/n}\,\mathrm{Vol}_n(B)^{1/n},
\end{aligned}$$

and it remains to notice that $S_{n-1}(B) = n\,\mathrm{Vol}_n(B)$. $\square$

The Prékopa–Leindler inequality represents the case $p = 0$ of the following general result:

Theorem 3.12.5 (The Borell–Brascamp–Lieb inequality) *Suppose that $0 < \lambda < 1$, $-1/n \leq p \leq \infty$, and f, g, and h are nonnegative integrable functions on $\mathbb{R}^n$ satisfying*

$$h((1-\lambda)x + \lambda y) \geq M_p(f(x), g(y)\,; 1-\lambda, \lambda),$$

for all $x, y \in \mathbb{R}^n$. Then

$$\int_{\mathbb{R}^n} h(x)\,dx \geq M_{p/(np+1)}\left(\int_{\mathbb{R}^n} f(x)\,dx, \int_{\mathbb{R}^n} g(x)\,dx\,; 1-\lambda, \lambda\right).$$

Here $p/(np+1)$ means $-\infty$, if $p = -1/n$, and $1/n$, if $p = \infty$.

Proof. We start with the case $n = 1$. Without loss of generality we may assume that

$$\int_{\mathbb{R}} f(x)\,dx = A > 0 \quad \text{and} \quad \int_{\mathbb{R}} g(x)\,dx = B > 0.$$

We define $u, v\colon [0,1] \to \mathbb{R}$ such that $u(t)$ and $v(t)$ are the smallest numbers satisfying

$$\frac{1}{A}\int_{-\infty}^{u(t)} f(x)\,dx = \frac{1}{B}\int_{-\infty}^{v(t)} g(x)\,dx = t.$$

Clearly, the two functions are increasing and thus they are differentiable almost everywhere. This yields

$$\frac{f(u(t))u'(t)}{A} = \frac{g(v(t))v'(t)}{B} = 1 \quad \text{almost everywhere,}$$

so that $w(t) = (1-\lambda)u(t) + \lambda v(t)$ verifies

$$\begin{aligned} w'(t) &= (1-\lambda)u'(t) + \lambda v'(t) \\ &= (1-\lambda)\frac{A}{f(u(t))} + \lambda \frac{B}{g(v(t))} \end{aligned}$$

at every t with $f(u(t)) > 0$ and $g(v(t)) > 0$. Or,

$$\int_{\mathbb{R}} h(x)\, dx \geq \int_0^1 h(w(t))w'(t)\, dt$$

and the last inequality can be continued as

$$\begin{aligned} &\geq \int_0^1 M_p\big(f(u(t)), g(v(t))\,; 1-\lambda, \lambda\big) M_1\Big(\frac{A}{f(u(t))}, \frac{B}{g(v(t))}\,; 1-\lambda, \lambda\Big)\, dt \\ &\geq \int_0^1 M_{p/(p+1)}(A, B\,; 1-\lambda, \lambda)\, dt \\ &= M_{p/(p+1)}(A, B\,; 1-\lambda, \lambda), \end{aligned}$$

by a generalization of the discrete Rogers–Hölder inequality (provided by Section 1.2, Exercise 4 (i), for $n = 2$, $q = 1$, and $p + q \geq 0$).

The general case follows by induction. Suppose that it is true for all natural numbers less than n.

For each $s \in \mathbb{R}$, attach to f, g, and h section functions f_s, g_s, and h_s, following the model

$$f_s \colon \mathbb{R}^{n-1} \to \mathbb{R}, \quad f_s(z) = f(z, s).$$

Let $x, y \in \mathbb{R}^{n-1}$, let $a, b \in \mathbb{R}$ and put $c = (1-\lambda)a + \lambda b$. Then

$$\begin{aligned} h_c((1-\lambda)x + \lambda y) &= h((1-\lambda)x + \lambda y, (1-\lambda)a + \lambda b) \\ &= h((1-\lambda)(x, a) + \lambda(y, b)) \\ &\geq M_p(f(x, a), g(y, b)\,; 1-\lambda, \lambda) \\ &= M_p(f_a(x), g_b(y)\,; 1-\lambda, \lambda) \end{aligned}$$

and thus, by our inductive hypothesis,

$$\int_{\mathbb{R}^{n-1}} h_c(x)\, dx \geq M_{p/((n-1)p+1)}\Big(\int_{\mathbb{R}^{n-1}} f_a(x)\, dx, \int_{\mathbb{R}^{n-1}} g_b(x)\, dx\,; 1-\lambda, \lambda\Big).$$

Letting

$$H(c)=\int_{\mathbb{R}^{n-1}} h_c(x)\,dx, \quad F(a)=\int_{\mathbb{R}^{n-1}} f_a(x)\,dx, \quad G(b)=\int_{\mathbb{R}^{n-1}} g_b(x)\,dx,$$

we have

$$H(c)=H((1-\lambda)a+\lambda b)\geq M_r(F(a),G(b)\,;1-\lambda,\lambda),$$

where $r=p/((n-1)p+1)$, so by Fubini's theorem and our inductive hypothesis we conclude that

$$\begin{aligned}\int_{\mathbb{R}^n} h(x)\,dx &\geq \int_{\mathbb{R}}\int_{\mathbb{R}^{n-1}} h_c(z)\,dz\,dc=\int_{\mathbb{R}} H(c)\,dc\\ &\geq M_{r/(r+1)}\left(\int_{\mathbb{R}} F(a)\,da,\int_{\mathbb{R}} G(b)\,db\,;1-\lambda,\lambda\right)\\ &= M_{p/(np+1)}\left(\int_{\mathbb{R}^n} f(x)\,dx,\int_{\mathbb{R}^n} g(x)\,dx\,;1-\lambda,\lambda\right).\end{aligned}$$

$\square$

The above argument of Theorem 3.12.5 goes back to R. Henstock and A. M. Macbeath [101] (when $n=1$) and illustrates a powerful tool of convex analysis: the Brenier map. See the Comments at the end of this chapter. Basically the same argument (plus some computation that make the objective of Exercise 10) led F. Barthe [15], [16] to a simplified approach of the best constants in some famous inequalities like the Young inequality and the reverse Young inequality.

Exercises

1. Verify Minkowski's formula for the surface area of a convex body K in the following particular cases:

(i) K is a disc;
(ii) K is a rectangle;
(iii) K is a regular tetrahedron.

2. Infer from the isoperimetric inequality for convex bodies in $\mathbb{R}^n$ the following classical result: If A is the area of a domain in plane, bounded by a curve of length L, then

$$L^2\geq 4\pi A$$

and the equality holds only for discs.

3. Settle the equality case in the Brunn–Minkowski inequality (as stated in Theorem 3.12.3).

Remark. The equality case in the Prékopa–Leindler inequality is open.

4. Let $f = f(x,y)$ be an integrable log-concave function defined on an open convex set Ω of $\mathbb{R}^{m+n}$ and for each x in the orthogonal projection $\mathrm{pr}_1\,\Omega$, of Ω onto $\mathbb{R}^m$, define

$$F(x) = \int_{\Omega(x)} f(x,y)\,dy,$$

where $\Omega(x) = \{y \in \mathbb{R}^n \mid (x,y) \in \Omega\}$. Infer from the Prékopa–Leindler inequality that the function $F(x)$ is log-concave on $\mathrm{pr}_1\,\Omega$.
[*Hint*: Suppose that $x_k \in \mathrm{pr}_1\,\Omega$ and $y_k \in \Omega(x_k)$ for $k = 1,2$ and $\lambda \in (0,1)$. Then

$$\Omega((1-\lambda)x_1 + \lambda x_2) \supset (1-\lambda)\Omega(x_1) + \lambda\Omega(x_2)$$

and

$$f((1-\lambda)x_1 + \lambda x_2, (1-\lambda)y_1 + \lambda y_2) \geq f(x_1,y_1)^{1-\lambda} f(x_2,y_2)^{\lambda}. \quad]$$

5. Consider open convex sets Ω_k in $\mathbb{R}^n$ and log-concave functions f_k, from Ω_k into $\mathbb{R}_+$, where $k = 1,2$.

(i) Prove that the function $f(x,y) = f_1(x)f_2(y)$ is log-concave on the open convex set $\Omega_1 \times \Omega_2$.

(ii) Infer from (i) and Exercise 4 that the convolution $f_1 * f_2$ is log-concave on $\Omega_1 + \Omega_2$.

6. Consider a convex body K in $\mathbb{R}^3$. Infer from Exercise 4 that the square root of the area of the cross section of K by parallel hyperplanes is a concave function. See Fig. 3.3.

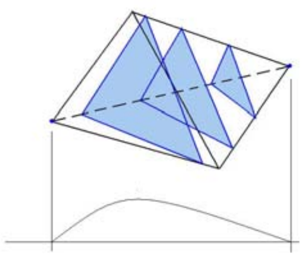

Fig. 3.3. Unimodal cross sections.

More precisely, if $v \in \mathbb{R}^3$ and $\|v\| \neq 0$, then the function

$$A(t) = \sqrt{S_{n-1}(K \cap \{x \mid \langle x, v\rangle = t\})}$$

is concave on the interval consisting of all t for which

$$K \cap \{x \mid \langle x, v\rangle = t\} \neq \emptyset.$$

7. (The essential form of the Prékopa–Leindler inequality; see H. J. Brascamp and E. H. Lieb [39]) Let $f, g \in L^1(\mathbb{R}^n)$ be two nonnegative functions and let $\lambda \in (0,1)$. The function

$$S(x) = \operatorname*{ess\,sup}_{y} f\Big(\frac{x-y}{1-\lambda}\Big)^{1-\lambda} g\Big(\frac{y}{\lambda}\Big)^{\lambda}$$

is measurable since

$$S(x) = \sup_n \int_{\mathbb{R}^n} f\Big(\frac{x-y}{1-\lambda}\Big)^{1-\lambda} g\Big(\frac{y}{\lambda}\Big)^{\lambda} \varphi_n(y)\,dy$$

for every sequence $(\varphi_n)_n$, dense in the unit ball of $L^1(\mathbb{R}^n)$. Prove that

$$\|S\|_{L^1} \geq \|f\|_{L^1}^{1-\lambda}\,\|g\|_{L^1}^{\lambda} \tag{3.33}$$

and derive from this result the classical Prékopa–Leindler inequality.

Remark. As noticed in [39], the essential form of the Prékopa–Leindler inequality represents the limiting case as $r \to 0+$, of the following reverse Young inequality with sharp constants: *Let* $0 < p, q, r \leq 1$ *with* $1/p + 1/q = 1 + 1/r$, *and let* $f \in L^p(\mathbb{R}^n)$ *and* $g \in L^q(\mathbb{R}^n)$ *be nonnegative functions. Then*

$$\|f * g\|_{L^r} \geq C(p,q,r,n)\|f\|_{L^p}\,\|g\|_{L^q}.$$

8. A nonnegative regular measure μ defined on the (Lebesgue) measurable subsets of $\mathbb{R}^n$ is called M_p-*concave* (for some $p \in \overline{\mathbb{R}}$) if

$$\mu((1-\lambda)X + \lambda Y) \geq M_p(\mu(X), \mu(Y)\,; 1-\lambda, \lambda)$$

for all measurable sets X and Y in $\mathbb{R}^n$ and all $\lambda \in (0,1)$ such that the set $(1-\lambda)X + \lambda Y$ is measurable. When $p = 0$, a M_p-concave measure is also called log-*concave*. By the Prékopa–Leindler inequality, the Lebesgue measure is $M_{1/n}$-concave. Suppose that $-1/n \leq p \leq \infty$, and let f be a nonnegative integrable function which is M_p-concave on an open convex set C in $\mathbb{R}^n$. Prove that the measure $\mu(X) = \int_{C\cap X} f(x)\,dx$ is $M_{p/(np+1)}$-concave. Infer that the standard *Gauss measure* in $\mathbb{R}^n$,

$$d\gamma_n = (2\pi)^{-n/2}\mathrm{e}^{-\|x\|^2/2}dx,$$

is log-concave.

9. (S. Dancs and B. Uhrin [62]) Extend Theorem 3.12.5 by replacing the Lebesgue measure by a M_q-concave measure, for some $-\infty \leq q \leq 1/n$.

10. (F. Barthe [15]) Let $v_1, \dots, v_m$ be vectors in $\mathbb{R}^n$ $(m \geq n)$, and $c_1, \dots, c_m$ be positive numbers such that $\sum_{k=1}^m c_k = n$. For $\lambda = (\lambda_1, \dots, \lambda_m) \in (0,\infty)^m$, consider the following two norms on $\mathbb{R}^n$:

$$M_\lambda(x) = \inf\left\{\left(\sum_{k=1}^m c_k\theta_k^2/\lambda_k\right)^{1/2} \mid x = \sum_{k=1}^m c_k\theta_k v_k,\ \theta_k \in \mathbb{R}\right\}$$

and

$$N_\lambda(x) = \left(\sum_{k=1}^m c_k\lambda_k\langle x, v_k\rangle^2\right)^{1/2}$$

Prove that $\mathcal{E}_\lambda = \{x \in \mathbb{R}^n \mid M_\lambda(x) \le 1\}$ is the polar of the ellipsoid $\mathcal{F}_\lambda = \{x \in \mathbb{R}^n \mid N_\lambda(x) \le 1\}$ and

$$\mathrm{Vol}_n(\mathcal{E}_\lambda)\,\mathrm{Vol}_n(\mathcal{F}_\lambda) = \mathrm{Vol}_n(B)^2,$$

where B denotes the closed unit ball of the Euclidean space $\mathbb{R}^n$.
[*Hint*: Notice that the support function of the polar of $\mathcal{F}_\lambda$, is

$$h(\mathcal{F}_\lambda^\circ)(x) = \left(\sum_{k=1}^m c_k\lambda_k\langle x, v_k\rangle^2\right)^{1/2}$$

which equals the support function of $\mathcal{E}_\lambda$,

$$h(\mathcal{E}_\lambda)(x) = \sup\left\{\sum_{k=1}^m c_k\theta_k\langle x, v_k\rangle \mid \sum_{k=1}^m c_k\theta_k^2/\lambda_k \le 1,\ \theta_k \in \mathbb{R}\right\}.\ ]$$

3.13 Mazur–Ulam Spaces and Convexity

Let E be a real normed linear space. The classical Mazur–Ulam theorem asserts that every bijective isometry $T\colon E \to E$ is an affine map, that is,

$$T(\lambda x + (1-\lambda)y) = \lambda T(x) + (1-\lambda)T(y) \tag{3.34}$$

for all $x, y \in E$ and $\lambda \in \mathbb{R}$. The essence of this result is the property of T to preserve midpoints of line segments, that is,

$$T\Big(\frac{x+y}{2}\Big) = \frac{T(x)+T(y)}{2} \tag{3.35}$$

for all $x, y \in E$. In fact, the condition (3.35) implies (3.34) for dyadic affine combinations, and thus for all convex combinations (since every isometry is a continuous map). Finally, it is routine to pass from convex combinations to general affine combinations in (3.34).

Surprisingly, the linear structure of E is needed only to support the notion of midpoint. In fact, a property like (3.35), of midpoint preservation, works in the framework of metric spaces as long as a well-behaved concept of midpoint is available. This is made clear by the following definition:

Definition 3.13.1 A *Mazur–Ulam space* is any metric space $M = (M, d)$ on which there is given a pairing $\sharp\colon M \times M \to M$ with the following four properties:

- (the idempotent property) $x \sharp x = x$ for all $x \in M$;
- (the commutative property) $x \sharp y = y \sharp x$ for all $x, y \in M$;
- (the midpoint property) $d(x, y) = 2d(x, x \sharp y) = 2d(y, x \sharp y)$ for all $x, y \in M$;
- (the transformation property) $T(x \sharp y) = T(x) \sharp T(y)$, for all $x, y \in M$ and all bijective isometries $T\colon M \to M$.

A Mazur–Ulam space should be viewed as a triplet $(M, d, \sharp)$. In this context, the point $x \sharp y$ is called a *midpoint* between x and y.

In a real normed space, the midpoint has the classical definition,

$$x \sharp y = \frac{x + y}{2},$$

and the Mazur–Ulam theorem is equivalent to the assertion that *every real normed space is a Mazur–Ulam space.* It is exactly in this way we want to extend the Mazur–Ulam theorem, by investigating other classes of Mazur–Ulam spaces, of which there are many.

In the above example, $\sharp$ coincides with the arithmetic mean, A. The simplest example of a Mazur–Ulam space where the midpoint is associated to the geometric mean is $M = (0, \infty)$, endowed with the metric

$$\delta(x, y) = \left|\log \frac{x}{y}\right|,$$

and the midpoint pairing $x \sharp y = G(x, y) = \sqrt{xy}$.

The Mazur–Ulam theorem can be proved easily by noticing the presence of sufficiently many reflections on any normed vector space. This idea can be considerably extended.

Theorem 3.13.2 *Suppose that $M = (M, d)$ is a metric space such that for every pair (a, b) of points of M there exists a bijective isometry $G_{(a,b)}$, from M onto itself, having the following two properties:*

(MU 1) $G_{(a,b)}a = b$ *and* $G_{(a,b)}b = a$;

(MU 2) $G_{(a,b)}$ *has a unique fixed point z (denoted $a \sharp b$) and*

$$d(G_{(a,b)}x, x) = 2d(x, z) \quad \text{for all } x \in M.$$

Then M is a Mazur–Ulam space.

The geometrical framework of Theorem 3.13.2 is illustrated in Fig. 3.4, while its proof will constitute the objective of Lemma 3.13.4 below.

Every normed vector space verifies the hypotheses of Theorem 3.13.2. In fact, in that case the maps $G_{(a,b)}$ are precisely the reflections

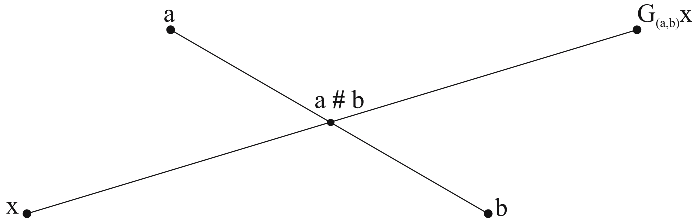

Fig. 3.4. The geometrical framework of Theorem 3.13.2.

$$G_{(a,b)}x = a + b - x.$$

The unique fixed point of $G_{(a,b)}$ is the midpoint of the line segment $[a, b]$, that is, $a \sharp b = (a + b)/2$.

In the case of $M = (\mathbb{R}_+^\star, \delta, G)$, the hypotheses of Theorem 3.13.2 are fulfilled by the family of isometries

$$G_{(a,b)}x = \frac{ab}{x}\,;$$

the fixed point of $G_{(a,b)}$ is precisely the geometric mean $\sqrt{ab}$, of a and b. A higher dimensional generalization of this example is provided by the space $\mathrm{Sym}^{++}(n, \mathbb{R})$, endowed with the *trace metric*,

$$d_{\mathrm{trace}}(A, B) = \Big(\sum_{k=1}^{n} \log^2 \lambda_k\Big)^{1/2}, \tag{3.36}$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of AB^{-1}. Since similarities preserve eigenvalues, this metric is invariant under similarities, that is,

$$d_{\mathrm{trace}}(A, B) = d_{\mathrm{trace}}(C^{-1}AC, C^{-1}BC) \quad \text{for all } C \in \mathrm{GL}(n, \mathbb{R}).$$

Note that AB^{-1} is similar with

$$A^{-1/2}(AB^{-1})A^{1/2} = A^{1/2}B^{-1/2}(A^{1/2}B^{-1/2})^\star > 0$$

and this fact assures the positivity of the eigenvalues of AB^{-1}.

The proof that $\mathrm{Sym}^{++}(n, \mathbb{R})$ admits a midpoint pairing follows from Theorem 3.13.2. We shall need the following technical result:

Lemma 3.13.3 *Given two matrices A and B in $\mathrm{Sym}^{++}(n, \mathbb{R})$, their geometric mean*

$$A \sharp B = A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2}$$

is the unique matrix C in $\mathrm{Sym}^{++}(n, \mathbb{R})$ such that

$$d_{\mathrm{trace}}(A, C) = d_{\mathrm{trace}}(B, C) = \frac{1}{2}\, d_{\mathrm{trace}}(A, B).$$

The geometric mean $A \sharp B$ of two positive definite matrices A and B was introduced by Pusz and Woronowicz [207]. It is the unique solution of the equation

$$XA^{-1}X = B$$

and this fact has a number of useful consequences such as:

$$A \sharp B = (AB)^{1/2} \quad \text{if } A \text{ and } B \text{ commute}$$
$$A \sharp B = B \sharp A$$
$$(C^*AC) \sharp (C^*BC) = C^*(A \sharp B)C \quad \text{for all } C \in \mathrm{GL}(n, \mathbb{R})$$

as well as the fact that the maps

$$G_{(A,B)}X = (A \sharp B)X^{-1}(A \sharp B)$$

verify the condition (MU 1) in Theorem 3.13.2 above. As concerns the condition (MU 2), let us check first the fixed points of $G_{(A,B)}$. Clearly, $A \sharp B$ is a fixed point. It is the only fixed point because any solution $X \in \mathrm{Sym}^{++}(n, \mathbb{R})$ of the equation

$$CX^{-1}C = X,$$

with $C \in \mathrm{Sym}^{++}(n, \mathbb{R})$, verifies the relation

$$(X^{-1/2}CX^{-1/2})(X^{-1/2}CX^{-1/2}) = I.$$

Since the square root is unique, we get $X^{-1/2}CX^{-1/2} = I$, that is, $X = C$. The second part of the condition (MU 2) asks for

$$d_{\mathrm{trace}}(G_{(A,B)}X, X) = 2d(X, A \sharp B),$$

that is,

$$d_{\mathrm{trace}}((A \sharp B)X^{-1}(A \sharp B), X) = 2d_{\mathrm{trace}}(X, A \sharp B),$$

for every $X \in \mathrm{Sym}^{++}(n, \mathbb{R})$. This follows directly from the definition (3.36) of the trace metric. Notice that $\sigma(C^2) = \{\lambda^2 \mid \lambda \in \sigma(C)\}$ for all C in $\mathrm{Sym}^{++}(n, \mathbb{R})$.

Lemma 3.13.4 *Suppose that $M_1 = (M_1, d_1)$ and $M_2 = (M_2, d_2)$ are two metric spaces which verify the conditions* (MU 1) *and* (MU 2) *of Theorem* 3.13.2. *Then*

$$T(x \sharp y) = Tx \sharp Ty$$

for all bijective isometries $T \colon M_1 \to M_2$, and all $x, y \in M$.

Proof. For $x, y \in M_1$ arbitrarily fixed, consider the set $\mathcal{G}_{(x,y)}$ of all bijective isometries $G \colon M_1 \to M_1$ such that $Gx = x$ and $Gy = y$. Notice that the identity of M_1 belongs to $\mathcal{G}_{(x,y)}$. Put

$$\alpha = \sup_{G \in \mathcal{G}_{(x,y)}} d(Gz, z),$$

where $z = x \sharp y$. Since

$$d(Gz, z) \le d(Gz, x) + d(x, z) = d(Gz, Gx) + d(x, z) = 2d(x, z),$$

we infer that $\alpha < \infty$. If $G \in \mathcal{G}_{(x,y)}$, so is $G' = G_{x,y}G^{-1}G_{x,y}G$, which yields

$$d(G_{(x,y)}G^{-1}G_{(x,y)}Gz, z) \le \alpha.$$

Then

$$\begin{aligned} d(G'z, z) &= d(G_{(x,y)}G^{-1}G_{(x,y)}Gz, z) = d(G_{(x,y)}G^{-1}G_{(x,y)}Gz, G_{x,y}z) \\ &= d(G^{-1}G_{(x,y)}Gz, z) \\ &= d(G_{(x,y)}Gz, Gz) \\ &= 2d(Gz, z) \end{aligned}$$

and thus $d(Gz, z) \le \alpha/2$ for all G. Consequently $\alpha = 0$ and this yields $G(z) = z$ for all $G \in \mathcal{G}_{(x,y)}$.

Now, for $T\colon M_1 \to M_2$ an arbitrary bijective isometry, we want to show that $Tz = z'$, where $z' = Tx \sharp Ty$. In fact, $G_{(x,y)}T^{-1}G_{(Tx,Ty)}T$ is a bijective isometry in $\mathcal{G}_{(x,y)}$, so

$$G_{(x,y)}T^{-1}G_{(Tx,Ty)}Tz = z.$$

This implies

$$G_{(Tx,Ty)}Tz = Tz.$$

Since z' is the only fixed point of $G_{(Tx,Ty)}$, we conclude that $Tz = z'$. $\square$

As observed by A. Vogt [241], the Mazur–Ulam theorem can be extended to all surjective maps $T\colon E \to F$ (acting on real normed spaces of dimension ≥ 2) which preserve equality of distances,

$$\|x - y\| = \|u - v\| \text{ implies } \|Tx - Ty\| = \|Tu - Tv\|.$$

It is open whether this result remains valid in the more general framework of Theorem 3.13.2.

The Mazur–Ulam spaces constitute a natural framework for a generalized theory of convexity, where the role of the arithmetic mean is played by a midpoint pairing.

Suppose that $M' = (M', d', \sharp')$ and $M'' = (M'', d'', \sharp'')$ are two Mazur–Ulam spaces, with M'' a subinterval of $\mathbb{R}$. A continuous function $f\colon M' \to M''$ is called *convex* (more precisely, $(\sharp', \sharp'')$-*convex*) if

$$f(x \sharp' y) \le f(x) \sharp'' f(y) \quad \text{for all } x, y \in M' \tag{3.37}$$

and *concave* if the opposite inequality holds. If

$$f(x \sharp' y) = f(x) \sharp'' f(y) \quad \text{for all } x, y \in M' \tag{3.38}$$

then the function f is called *affine.*

Every subinterval of $\mathbb{R}$ (endowed with the pairing associated to the arithmetic mean) is a Mazur–Ulam space and thus the above framework provides a generalization of the usual notion of convex function.

When $M' = M'' = (\mathbb{R}_+^\star, \delta, G)$, the convex functions $f\colon \mathbb{R}_+^\star \to \mathbb{R}_+^\star$ in the sense of (3.37) are precisely the multiplicatively convex functions. When $M' = \mathbb{R}$ and $M'' = (\mathbb{R}_+^\star, \delta, G)$, we recover the class of log-convex functions.

Things become considerably more technical when $\mathbb{R}_+^\star$ is replaced by the cone $\mathrm{Sym}^{++}(n, \mathbb{R})$. Here a number of basic questions are still open, for example, the generality of the Jensen inequality. The analogue of $(1-\lambda)x + \lambda y$ in the context of $\mathrm{Sym}^{++}(n, \mathbb{R})$ is

$$A \,\sharp_\lambda B = A^{1/2}(A^{-1/2}BA^{-1/2})^\lambda A^{1/2},$$

and this formula was investigated by F. Kubo and T. Ando [134] from the point of view of noncommutative means. What is the analogue of a convex combination for three (or finitely many) positive matrices? An interesting approach was recently proposed by T. Ando, C.-K. Li and R. Mathias [8], but the corresponding theory of convexity is still in its infancy.

Exercises

1. (The noncommutative analogue of two basic inequalities) The functional calculus with positive elements in $\mathfrak{A} = \mathrm{M}_n(\mathbb{R})$ immediately yields the following generalization of Bernoulli's inequality:

$$\begin{aligned} A^\alpha &\geq I + \alpha(A - I) \quad \text{if } \alpha \in (-\infty, 0] \cup [1, \infty) \\ A^\alpha &\leq I + \alpha(A - I) \quad \text{if } \alpha \in [0, 1], \end{aligned}$$

for all $A \in \mathfrak{A}$, $A > 0$. Infer that

$$\begin{aligned} A \,\sharp_\alpha\, B &\geq (1 - a)A + \alpha B \quad \text{if } \alpha \in (-\infty, 0] \cup [1, \infty) \\ A \,\sharp_\alpha\, B &\leq (1 - a)A + \alpha B \quad \text{if } \alpha \in [0, 1], \end{aligned}$$

for all $A, B > 0$ (which represents a generalization of Young's inequality).

Remark. The same argument works in the general framework of C^*-algebras. See [58] for details concerning these spaces.

2. (A generalization of the Rogers–Hölder inequality) Let $\mathfrak{A}$ be as in the preceding exercise and let $\varphi\colon \mathfrak{A} \to \mathbb{C}$ be a linear functional such that $\varphi(A) \geq 0$ if $A \geq 0$.

(i) Prove that

$$\varphi(A \,\sharp_\alpha\, B) \leq \varphi(A)^{1-\alpha}\varphi(B)^\alpha$$

for all $A, B \in \mathfrak{A}$, with $A, B > 0$ and all $\alpha \in [0, 1]$. This inequality works in the range $(-\infty, 0] \cup [1, \infty)$ with opposite inequality sign.

(ii) Letting $\varphi(A) = \langle Ax, x\rangle$ for some unit vector x, infer that

$$\langle A^{\alpha}x, x\rangle^{1/\alpha} \leq \langle Ax, x\rangle \quad \text{for all } A \in \mathfrak{A}, A \geq 0 \quad \text{and all } \alpha \in [0,1].$$

3.14 Comments

The first modern exposition on convexity in $\mathbb{R}^n$ was written by W. Fenchel [79]. He used the framework of lower semicontinuous proper convex functions to provide a valuable extension of the classical theory.

L. N. H. Bunt proved Theorem 3.2.2 in his Ph.D. thesis (1934). His priority, as well as the present status of Klee's problem, are described in a paper by J.-B. Hiriart-Urruty [104].

All the results in Section 3.3 on hyperplanes and separation theorems in $\mathbb{R}^n$ are due to H. Minkowski [164]. Their extension to the general context of linear topological spaces is presented in Appendix A.

Support functions, originally defined by H. Minkowski in the case of bounded convex sets, have been studied for general convex sets in $\mathbb{R}^n$ by W. Fenchel [78], [79], and in infinite-dimensional spaces by L. Hörmander [107].

The critical role played by finite dimensionality in a number of important results on convex functions is discussed by J. M. Borwein and A. S. Lewis in [38, Chapter 9].

A Banach space E is called *smooth* if at each point of its unit sphere there is a unique hyperplane of support for the closed unit ball. Equivalently, E is smooth if and only if the norm function is Gâteaux differentiable at every $x \neq 0$. In the context of separable Banach spaces, one can prove that the points of the unit sphere S where the norm is Gâteaux differentiable form a countable intersection of dense open subsets of S (and thus they constitute a dense subset, according to the Baire category theorem). See [200, p. 43]. The book by M. M. Day [64] contains a good account on the problem of renorming Banach spaces to improve the smoothness properties.

A Banach space E is said to be a *weak* (*strong*) *differentiability space* if for each convex open set U in E and each continuous convex function $f\colon U \to \mathbb{R}$ the set of points of Gâteaux (Fréchet) differentiability of f contains a dense G_δ subset of E. E. Asplund [11], indicated rather general conditions under which a Banach space has a renorming with this property. See R. R. Phelps [199] for a survey on the differentiability properties of convex functions on a Banach space.

The convex functions can be characterized in terms of distributional derivatives: *If Ω is an open convex subset of $\mathbb{R}^n$, and $f\colon \Omega \to \mathbb{R}$ is a convex function, then Df is monotone, and $D^2 f$ is a positive and symmetric* (*matrix-valued and locally bounded*) *measure. Conversely, if f is locally integrable and $D^2 f$ is a positive* (*matrix-valued*) *distribution on Ω, then f agrees* *almost everywhere on Ω with a convex function g such that $\Omega \subset \operatorname{dom} g$.* See [4, Proposition 7.11].

F. Mignot [163] proved the following generalization of Rademacher's theorem: *Every Lipschitz function from a separable Hilbert space into a Hilbert space is Gâteaux differentiable at densely many points.* However, as shown in the case of the function

$$F\colon L^2[0,1] \to \mathbb{R}, \quad F(f) = \left(\int_0^1 (f^+(t))^2 \, dt \right)^{1/2},$$

the set of points of Fréchet differentiability may be empty. The hypothesis on separability is essential for the validity of Mignot's result. A counterexample is provided by the projection of $\ell^2(I)$ (for an uncountable index set I) onto the cone of positive elements.

Proofs of Alexandrov's theorem (Theorem 3.11.2) may be found in G. Alberti and L. Ambrosio [4], G. Bianchi, A. Colesanti and C. Pucci [28] (which includes also some historical comments), M. G. Crandall, H. Ishii and P.-L. Lions [61], L. C. Evans and R. Gariepy [74], and F. Mignot [163]. The proof in our text follows [61]. In Mignot's approach, Alexandrov's theorem appears as a consequence of the following differentiability property of monotone maps:

Differentiability of monotone maps. *Let u be a maximal monotone map on $\mathbb{R}^n$ and let D be the set of points x such that $u(x)$ is a singleton. Then u is differentiable at almost every $a \in D$, that is, there exists an $n \times n$ matrix $\nabla u(a)$ such that*

$$\lim_{\substack{x \to a \\ y \in u(x)}} \frac{y - u(a) - \nabla u(a)(x-a)}{\|x-a\|} = 0.$$

In fact, if we apply this result to the subdifferential of a convex function $f\colon \mathbb{R}^n \to \mathbb{R}$, we obtain that for almost every $a \in \mathbb{R}^n$ where ∂f is a singleton (that is, where f is differentiable), there exists a matrix $\nabla^2 f(a)$ such that

$$\lim_{\substack{x \to a \\ y \in \partial f(x)}} \frac{y - \nabla f(a) - \nabla^2 f(a)(x-a)}{\|x-a\|} = 0. \tag{3.39}$$

If (3.39) holds, then $A = \nabla^2 f(a)$ proves to be the Alexandrov Hessian of f at a. To show this, it suffices to restrict ourselves to the case where $a = 0$, $f(a) = 0$ and $\nabla f(a) = 0$. We shall prove that $\varphi(h) = f(h) - \frac{1}{2}\langle Ah, h\rangle$ verifies $\lim_{h\to 0} \varphi(h)/\|h\|^2 = 0$. In fact, fixing an $h \neq 0$, by the nonsmooth version of the mean value theorem (see [56, Theorem 2.3.7]) we get a point y in the segment joining 0 to h, and a $p \in \partial\varphi(y)$ such that $\varphi(h) - \varphi(0) = \langle h, p\rangle$. Then

$$\varphi(h) = \langle q - Ay, h\rangle$$

for some vector $q \in \partial f(y)$. According to (3.39), $\lim_{h\to 0} \|q - A(y)\|/\|h\| = 0$, which yields $\lim_{h\to 0} \varphi(h)/\|h\|^2 = 0$.

Important applications of the infimal convolution and Alexandrov's theorem to the theory of viscosity solutions are presented in the remarkable user's guide written by M. G. Crandall, H. Ishii and P.-L. Lions [61].

A survey of the convexity properties of solutions of partial differential equations may be found in the notes of B. Kawohl [122]. We shall mention here one application of the Prékopa–Leindler inequality which refers to the diffusion equation

$$\frac{\partial u}{\partial t} = \frac{1}{2}\,\Delta u - V(x)u \quad \text{for } (x,t) \in \Omega \times (0,\infty)$$

with zero Dirichlet boundary conditions (that is, $\lim_{x\to\partial\Omega} u(x,t) = 0$ for each t). Here Ω is an open convex set in $\mathbb{R}^n$ and V is a nonnegative continuous function defined on Ω. When $\Omega = \mathbb{R}^n$ and $V = 0$, the fundamental solution is given by formula $f(x,y,t) = (2\pi t)^{-n/2}\mathrm{e}^{-\|x-y\|^2/2t}$, which is log-concave on $\mathbb{R}^n \times \mathbb{R}^n$. H. J. Brascamp and E. H. Lieb [40] have proved, based on the Prékopa–Leindler inequality, that in general the fundamental solution $f(x,y,t)$ of the above Dirichlet problem is log-concave on $\Omega\times\Omega$, whenever V is a convex function. The idea is to show that $f(x,y,t)$ is a pointwise limit of convolutions of log-concave functions (in which case Exercise 5, Section 3.12, applies). Later on, Ch. Borell [34] considered potentials $V = V(x,\sigma)$ that depend on a parameter, and this fact led him to more general results and a Brownian motion treatment of the Brunn–Minkowski inequality.

The Borell–Brascamp–Lieb inequality was first stated and proved in full generality by Ch. Borell [33] and H. J. Brascamp and E. H. Lieb [40]. Our presentation in Section 3.12 left untouched many important applications and ramifications. Fortunately, they are covered in a remarkable paper by R. J. Gardner [88] (nicely complemented by its electronic version [87]).

The Prékopa–Leindler inequality is related to optimal mass transport (which in turn provides a powerful technique to derive a number of interesting inequalities). Let μ and ν be two Borel probability measures on $\mathbb{R}^n$. A map $T\colon \mathbb{R}^n \to \mathbb{R}^n$ (defined μ-almost everywhere) is said to *push μ forward to ν* (or to *transport μ onto ν*) if $\nu(B) = \mu(T^{-1}(B))$ for every Borel set B in $\mathbb{R}^n$. Y. Brenier [41] found a very special map pushing forward one probability to another. His result was reconsidered by R. J. McCann [161], [162], who noticed that the absolutely continuous Borel probability measures can be transported by maps of the form $T = \nabla\varphi$, where φ is convex. These maps are usually referred to as *Brenier maps*. The differentiability properties of T (motivated by the existence of the Alexandrov Hessian of φ) makes possible an easy handling of T. For example, if

$$\mu(B) = \int_B f(x)\,dx \quad \text{and} \quad \nu(B) = \int_B g(x)\,dx$$

and $T = \nabla\varphi$ is the Brenier map (pushing μ forward to ν), then

$$\int_{\mathbb{R}^n} h(y)g(y)\,dy = \int_{\mathbb{R}^n} h(\nabla\varphi(x))f(x)\,dx$$

for all bounded Borel functions $h\colon \mathbb{R} \to \mathbb{R}_+$. Assuming the change of variable $y = \nabla\varphi(x)$ is working, the last formula leads to the so-called *Monge–Ampère equation*,

$$f(x) = \det(\operatorname{Hess}_x \varphi) \cdot g(\nabla\varphi(x)). \tag{3.40}$$

As noted by R. J. McCann [162], this equation is valid in general, provided that $\operatorname{Hess}_x \varphi$ is replaced by the Alexandrov Hessian of φ.

When $n = 1$, then we can find a T that transports μ to ν by defining $T(t)$ to be the smallest number such that

$$\int_{-\infty}^{t} f(x)\,dx = \int_{-\infty}^{T(t)} g(x)\,dx.$$

This is the key parametrization in proving the Prékopa–Leindler inequality (first noticed in this form by R. Henstock and A. M. Macbeath [101]).

Applications of the mass transport theory may be found in the recent book of C. Villani [240] (see also [87] and [88]). A sample is the following classical result (due to L. Gross in the case where μ is the standard Gauss measure $d\gamma_n = (1/\sqrt{2\pi})^n \mathrm{e}^{-\|x\|^2/2} dx$), for which D. Cordero-Erausquin [59] has found a simple argument based on the Brenier maps:

The logarithmic Sobolev inequality. *Let μ be a probability measure on $\mathbb{R}^n$ of the form $d\mu = \mathrm{e}^{-V(x)}dx$, where V is a twice differentiable function satisfying $\operatorname{Hess}_x V \geq cI$ for some $c > 0$. Then, for every smooth nonnegative function f on $\mathbb{R}^n$,*

$$\operatorname{Ent}_\mu(f) \leq \frac{1}{2c}\, I_\mu(f).$$

Here

$$\operatorname{Ent}_\mu(f) = \int_{\mathbb{R}^n} f \log f\, d\mu - \left(\int_{\mathbb{R}^n} f\, d\mu\right)\left(\int_{\mathbb{R}^n} \log f\, d\mu\right)$$

represents the *entropy* of f and

$$I_\mu(f) = \int_{\mathbb{R}^n} \frac{\|\nabla f\|^2}{f}\, d\mu$$

represents the *Fischer-information* of f.

D. Cordero-Erausquin's idea in the case of standard Gauss measures is both simple and instructive. Without loss of generality we may assume that $\int_{\mathbb{R}^n} f\, d\gamma_n = 1$. Denote by $\nabla\varphi$ the Brenier map that transports $f d\gamma_n$ into $d\gamma_n$ and put $\theta(x) = \varphi(x) - \|x\|^2/2$. Then

$$\nabla\varphi(x) = x + \nabla\theta(x) \quad \text{and} \quad I + \operatorname{Hess}_x \theta \geq 0.$$

The corresponding Monge–Ampère equation reads as

$$f(x)\mathrm{e}^{-\|x\|^2/2} = \det(I + \operatorname{Hess}_x \theta)\mathrm{e}^{-\|x+\nabla\theta(x)\|^2/2},$$

which yields (taking into account that $\log(1+t) \le t$ for $t > -1$),

$$\begin{aligned}\log f(x) &= -\|x + \nabla\theta(x)\|^2/2 + \|x\|^2/2 + \log\det(I + \mathrm{Hess}_x\,\theta)\\ &= -x\cdot\nabla\theta(x) - \|\nabla\theta(x)\|^2/2 + \log\det(I + \mathrm{Hess}_x\,\theta)\\ &\le -x\cdot\nabla\theta(x) - \|\nabla\theta(x)\|^2/2 + \Delta\theta(x).\end{aligned}$$

By integrating both sides with respect to $f d\gamma_n$ we infer that

$$\int_{\mathbb{R}^n} f\log f\,d\gamma_n \le \int_{\mathbb{R}^n} f[\Delta\theta - x\cdot\nabla\theta]\,d\gamma_n - \frac{1}{2}\int_{\mathbb{R}^n}\|\nabla\theta(x)\|^2\,d\gamma_n.$$

Now use Exercise 5, Section 3.11, to get

$$\int_{\mathbb{R}^n} f\Delta\theta\,d\gamma_n \le -\int_{\mathbb{R}^n}\nabla f\cdot\nabla\theta\,d\gamma_n.$$

This allows us to complete the proof as follows:

$$\begin{aligned}\int_{\mathbb{R}^n} f\log f\,d\gamma_n &\le -\int_{\mathbb{R}^n}\nabla f(x)\cdot\nabla\theta(x)\,d\gamma_n - \frac{1}{2}\int_{\mathbb{R}^n}\|\nabla\theta(x)\|^2\,d\gamma_n\\ &= -\frac{1}{2}\int_{\mathbb{R}^n}\left\|f^{1/2}\nabla\theta(x) + \frac{\nabla f(x)}{f^{1/2}}\right\|^2 d\gamma_n + \frac{1}{2}\int_{\mathbb{R}^n}\frac{\|\nabla f(x)\|^2}{f}\,d\gamma_n\\ &\le \frac{1}{2}\int_{\mathbb{R}^n}\frac{\|\nabla f(x)\|^2}{f}\,d\gamma_n = \frac{1}{2}\,I_{\gamma_n}(f),\end{aligned}$$

where the first equality is motivated by using integration by parts. □

The logarithmic Sobolev inequality (in various forms) has proved useful in several fields of mathematics. In PDE, it provides a control of the entropy production for evolutive dissipative systems. In probability theory, it is a tool to obtain concentration of measure phenomena or to study smoothness properties of Markov processes. In combinatorial theory, it gives estimates on mixing time of randomized algorithms. See C. Villani [240] and the references therein.

D. Cordero-Erausquin, R. J. McCann and M. Schmuckenschläger [60] have extended the Borell–Brascamp–Lieb inequality to the Riemannian setting. Let M be a complete, connected, n-dimensional manifold equipped with a Riemannian metric tensor given by C^2-smooth functions g_{ij} in coordinates. The role of $(1-t)x + ty$ is played by

$$Z_t(x,y) = \{z \in M \mid d(x,z) = td(x,y) \text{ and } d(z,y) = (1-t)d(x,y)\}.$$

Put $Z_t(x,Y) = \bigcup_{y\in Y} Z_t(x,y)$ and

$$v_t(x,y) = \lim_{r\to 0}\frac{\mathrm{Vol}_n[Z_t(x,B_r(y))]}{\mathrm{Vol}_n[B_{tr}(y)]}.$$

This ratio measures the volume distortion due to the curvature. In Euclidean space, $v_t(x,y) = 1$.

The Riemannian Borell–Brascamp–Lieb inequality. *Let f, g, h be nonnegative functions on M and let A, B be Borel subsets of M such that*

$$\int_A f\,dV = \int_B g\,dV = 1,$$

where dV denotes the volume measure on M. Assume that for all (x,y) in $A \times B$ and all z in $Z_t(x,y)$ we have

$$1/h(z)^{1/n} \leq (1-t)\Big[\frac{v_{1-t}(y,x)}{f(x)}\Big]^{1/n} + t\Big[\frac{v_t(x,y)}{g(y)}\Big]^{1/n}.$$

Then $\int_{\mathbb{R}^n} h\,dV \geq 1$.

The Mazur–Ulam theorem appeared in [160]. The concept of a Mazur–Ulam space was introduced by C. P. Niculescu [184], inspired by a recent argument given by J. Väisälä [239] to the Mazur–Ulam theorem, and also by a paper of J. D. Lawson and Y. Lim [140] on the geometric mean in the noncommutative setting. The presence of $\mathrm{Sym}^{++}(n, \mathbb{R})$ among the Mazur–Ulam spaces is just the tip of the iceberg. In fact, many other symmetric cones (related to the theory of Bruhat–Tits spaces in differential geometry) have the same property. See [184] and references therein.

The theory of convex functions of one real variable can be generalized to several variables in many different ways. A long time ago, P. Montel [171] pointed out the alternative to subharmonic functions. They are motivated by the fact that the higher analogue of the second derivative is the Laplacian. In a more recent paper, B. Kawohl [121] discussed the question when the superharmonic functions are concave. Nowadays, many other alternatives are known. An authoritative monograph on this subject has been published by L. Hörmander [108].

Linear programming is the mathematics of linear inequalities and thus it represents a natural generalization of linear algebra (which deals with linear equations). The theoretical basis of linear and nonlinear programming was published in 1902 by Julius Farkas, who gave a long proof of his result (Lemma 3.10.3 in our text).

4

Choquet's Theory and Beyond

The classical Hermite–Hadamard inequality, already presented in Section 1.9, gives us an estimate, from below and from above, of the arithmetic mean of a continuous convex function $f\colon [a,b] \to \mathbb{R}$:

$$f\Big(\frac{a+b}{2}\Big) \leq \frac{1}{b-a}\int_a^b f(x)\,dx \leq \frac{f(a)+f(b)}{2}.$$

Thinking of $[a,b]$ as a loaded bar with a uniform mass distribution, its barycenter is precisely the middle point $(a+b)/2$. In this setting the function f can be interpreted as a force density. When f is convex, the inequality above says that the arithmetic mean of f on $[a,b]$ lies between the value of f at the barycenter and the arithmetic mean of f at the endpoints. It is remarkable that this fact extends to all continuous convex functions defined on arbitrary compact convex sets. The details are covered by Choquet's theory, which forms the core of this chapter. This theory is quite demanding, and the reader should have some knowledge of the Krein–Milman theorem and the Riesz–Kakutani representation theory (along with the functional analysis and measure theory implicit in an understanding of these theorems). Both topics are presented in great detail in books like *Analysis II* by S. Lang [138]. For the convenience of the reader we have included in Appendix A the basic facts on the separation of convex sets in locally convex Hausdorff spaces, and also a proof of the Krein–Milman theorem. The background on measure theory can be covered from many sources. Besides [138], we mention here books by L. C. Evans and R. F. Gariepy [74], E. Hewitt and K. Stromberg [103] and W. Rudin [218].

4.1 Steffensen–Popoviciu Measures

Throughout this chapter K will denote a (nonempty) compact convex subset of a real locally convex Hausdorff space E, and $C(K)$ will denote the space of all real-valued continuous functions on K. We want to relate the geometry of K with the cone $\mathrm{Conv}(K)$, of all real-valued continuous convex functions defined on K.

According to the Stone–Weierstrass theorem, $\mathrm{Conv}(K)-\mathrm{Conv}(K)$ is dense in $C(K)$. In fact, due to the formula

$$\sup\{f_1 - g_1, f_2 - g_2\} = \sup\{f_1 + g_2, f_2 + g_1\} - (g_1 + g_2),$$

the set $\mathrm{Conv}(K) - \mathrm{Conv}(K)$ is a linear sublattice of $C(K)$ which contains the unit and separates the points of K (since $\mathrm{Conv}(K)$ contains all restrictions to K of the functionals $x' \in E'$).

We shall need also the space

$$A(K) = \mathrm{Conv}(K) \cap -\mathrm{Conv}(K),$$

of all real-valued continuous affine functions on K. This is a rich space, as the following result shows:

Lemma 4.1.1 $A(K)$ *contains*

$$E'|_K + \mathbb{R}\cdot 1 = \{x'|_K + \alpha \mid x' \in E' \text{ and } \alpha \in \mathbb{R}\}$$

as a dense subspace.

Proof. Let $f \in A(K)$ and $\varepsilon > 0$. The following two subsets of $E \times \mathbb{R}$,

$$J_1 = \{(x, f(x)) \mid x \in K\}$$

and

$$J_2 = \{(x, f(x) + \varepsilon) \mid x \in K\},$$

are nonempty, compact, convex and disjoint. By a geometric version of the Hahn–Banach theorem (see Theorem A.2.4), there exists a continuous linear functional L on $E \times \mathbb{R}$ and a number $\lambda \in \mathbb{R}$ such that

$$\sup L(J_1) < \lambda < \inf L(J_2).$$

Hence the equation $L(x, g(x)) = \lambda$ defines an element $g \in E'|_K + \mathbb{R}\cdot 1$ such that

$$f(x) < g(x) < f(x) + \varepsilon \quad \text{for all } x \in K.$$

In fact, $\lambda = L(x, 0) + g(x)L(0, 1)$, and thus $g(x) = (\lambda - L(x, 0))/L(0, 1)$. This solves the approximation (within ε) of f by elements of $E'|_K + \mathbb{R}\cdot 1$. $\square$

The following example shows that the inclusion $E'|_K + \mathbb{R}\cdot 1 \subset A(K)$ may be strict. For this, consider the set,

$$S = \{(a_n)_n \mid |a_n| \le 1/n^2 \text{ for every } n\},$$

viewed as a subset of ℓ^2 endowed with the weak topology. Then S is compact and convex and the function $f((a_n)_n) = \sum_n a_n$ defines an element of $A(S)$.

Moreover, $f(0) = 0$. However, there is no y in ℓ^2 such that $f(x) = \langle x, y\rangle$ for all $x \in S$.

The connection between the points of a compact convex set K and the positive functionals on $C(K)$ is made visible through the concept of barycenter. In physics, this concept is associated with material bodies. They can be thought of as compact sets S in $\mathbb{R}^3$ on which there is given a mass distribution, that is, a Borel measure μ with $\mu(S) > 0$. The barycenter is given by the formula

$$x_S = \frac{1}{\mu(S)} \int_S x \, d\mu(x), \tag{4.1}$$

and it is usual to say that x_S is the barycenter of μ rather then S. This leads to the notation x_μ for the barycenter.

In what follows we shall show that a concept of barycenter can be attached even to some signed Borel measures of positive total mass, defined on compact convex sets K (in a real locally convex Hausdorff space E).

Definition 4.1.2 A *Steffensen–Popoviciu measure* is any signed Borel measure μ on K such that

$$\mu(K) > 0 \quad \text{and} \quad \int_K f^+(x)\, d\mu(x) \geq 0 \quad \text{for all } f \in \mathrm{Conv}(K). \tag{4.2}$$

Clearly any Borel measure (of positive total mass) is also a Steffensen–Popoviciu measure. The following result provides a full characterization of these measures in the case of intervals.

Lemma 4.1.3 (T. Popoviciu [204]) *Let μ be a signed Borel measure on $[a, b]$ with $\mu([a, b]) > 0$. Then μ is a Steffensen–Popoviciu measure if and only if it verifies the following condition of end positivity,*

$$\int_a^t (t - x)\, d\mu(x) \geq 0 \quad \textit{and} \quad \int_t^b (x - t)\, d\mu(x) \geq 0, \tag{4.3}$$

for all $t \in [a, b]$.

Proof. Clearly, (4.2) yields

$$\int_a^b (x'(x) + t)^+ \, d\mu(x) \geq 0 \quad \text{for all } x' \in E' \ \text{ and all } t \in \mathbb{R}$$

and this is equivalent to (4.3) since the dual of $\mathbb{R}$ consists of the homotheties $x' \colon x \to sx$. The other implication, (4.3) $\Rightarrow$ (4.2), is based on Theorem 1.5.7. If $f \geq 0$ is a piecewise linear convex function, then f can be represented as a finite combination with nonnegative coefficients of functions of the form 1, $(x - t)^+$ and $(t - x)^+$, so that

$$\int_a^b f(x)\, d\mu(x) \geq 0.$$

The general case follows from this one, by approximating f^+ by piecewise linear convex functions. □

An alternative argument for (4.3) ⇒ (4.2), based on the integral representation of convex functions on intervals, was noticed by A. M. Fink [81].

Corollary 4.1.4 *Suppose that $x_1 \leq \cdots \leq x_n$ are real points and $p_1, \ldots, p_n$ are real weights. Then the discrete measure $\mu = \sum_{k=1}^n p_k \delta_{x_k}$ is a Steffensen–Popoviciu measure if and only if*

$$\sum_{k=1}^{n} p_k > 0, \quad \sum_{k=1}^{m} p_k(x_m - x_k) \geq 0 \quad \text{and} \quad \sum_{k=m}^{n} p_k(x_k - x_m) \geq 0 \tag{4.4}$$

for all $m \in \{1, \ldots, n\}$.

A special case when (4.4) holds is the following, used by Steffensen in his famous extension of Jensen's inequality (see Theorem 1.5.6):

$$\sum_{k=1}^{n} p_k > 0, \quad \text{and} \quad 0 \leq \sum_{k=1}^{m} p_k \leq \sum_{k=1}^{n} p_k, \quad \text{for all } m \in \{1, \ldots, n\}. \tag{4.5}$$

Corollary 4.1.5 *An absolutely continuous measure $d\mu = p(x)dx$ is a Steffensen–Popoviciu measure on $[a, b]$ if and only if*

$$\int_a^b p(x)\, dx > 0, \quad \int_a^t (t - x)p(x)\, dx \geq 0 \quad \text{and} \quad \int_t^b (x - t)p(x)\, dx \geq 0, \tag{4.6}$$

for all $t \in [a, b]$.

A stronger (but more suitable) condition than (4.6) is the following:

$$\int_a^b p(x)\, dx > 0 \quad \text{and} \quad 0 \leq \int_a^t p(x)\, dx \leq \int_a^b p(x)\, dx \quad \text{for all } t \in [a, b]. \tag{4.7}$$

As a consequence we obtain that $(x^2 + a)dx$ is a Steffensen–Popoviciu measure on $[-1, 1]$ for all $a > -1/3$. Notice that this is a signed measure if $a \in (-1/3, 0)$.

Integrating inequalities is not generally possible in the framework of signed measures. However, for the Steffensen–Popoviciu measures this works under certain restrictions, since (4.2) yields easily the following result:

Lemma 4.1.6 *Suppose that μ is a Steffensen–Popoviciu measure on K. If $h \in A(K)$, $f \in \mathrm{Conv}(K)$, and $h \leq f$, then*

$$\int_K h(x)\, d\mu(x) \leq \int_K f(x)\, d\mu(x).$$

An immediate consequence is as follows:

Corollary 4.1.7 *Suppose that μ is a Steffensen–Popoviciu measure on K and f is an affine function on K such that $\alpha \leq f \leq \beta$ for some real numbers α, β. Then*

$$\alpha \leq \frac{1}{\mu(K)} \int_K f(x)\, d\mu(x) \leq \beta.$$

According to the Riesz–Kakutani representation theorem (see [138], Theorem 9, p. 338), any continuous linear functional F on $C(K)$ can be uniquely represented by a signed Borel measure μ on K, via the formula

$$F(f) = \int_K f(x)\, d\mu(x), \quad f \in C(K).$$

Moreover, F is positive if and only if μ is positive. As a consequence, we can identify the linear functionals on a space $C(K)$ with their representative Borel measures.

By Corollary 4.1.7, if μ is a Steffensen–Popoviciu measure on K, then

$$\|\mu|_{A(K)}\| = \mu(K).$$

However, the norm of $\mu/\mu(K)$ as a functional on $C(K)$ can be arbitrarily large. In fact, in the case where $K = [-1, 1]$ and $d\mu = (x^2 + a)dx$, we have

$$\int_{-1}^{1} (x^2 + a)\, dx = \frac{2}{3} + 2a,$$

and thus

$$\left(\frac{2}{3} + 2a\right)^{-1} \int_{-1}^{1} |x^2 + a|\, dx = \frac{1}{1 + 3a}$$

for $a > -1/3$. This marks a serious difference from the case of positive Borel measures, where the norm of $\mu/\mu(K)$ is always 1.

Lemma 4.1.8 *Every Steffensen–Popoviciu measure μ on K admits a barycenter, that is, a point x_μ in K such that*

$$f(x_\mu) = \frac{1}{\mu(K)} \int_K f(x)\, d\mu(x) \tag{4.8}$$

for all continuous linear functionals f on E.

The barycenter x_μ is unique with this property. This is a consequence of the separability of the topology of E. See Corollary A.1.5. In the case of Euclidean spaces the norm and the weak convergence agree, so that the formula (4.8) reduces to (4.1).

Due to Lemma 4.1.1, the equality (4.8) extends to all $f \in A(K)$.

Proof of Lemma 4.1.8. We have to prove that

$$\Big(\bigcap_{f\in E'} H_f\Big)\cap K \neq \emptyset,$$

where H_f denotes the closed hyperplane $\{x \mid f(x) = \mu(f)/\mu(K)\}$ associated to $f \in E'$. Since K is compact, it suffices to prove that

$$\Big(\bigcap_{k=1}^{n} H_{f_k}\Big)\cap K \neq \emptyset$$

for every finite family $f_1, \ldots, f_n$ of functionals in E'. Equivalently, attaching to any such family of functionals the operator

$$T\colon K \to \mathbb{R}^n, \quad T(x) = (f_1(x), \ldots, f_n(x)),$$

we have to prove that $T(K)$ contains the point $p = \frac{1}{\mu(K)}(\mu(f_1), \ldots, \mu(f_n))$. In fact, if $p \notin T(K)$, then a separation argument yields an $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ such that

$$\langle p, a\rangle > \sup_{x\in K} \langle T(x), a\rangle,$$

that is,

$$\frac{1}{\mu(K)} \sum_{k=1}^{n} a_k \mu(f_k) > \sup_{x\in K} \sum_{k=1}^{n} a_k f_k(x).$$

Then $g = \sum_{k=1}^{n} a_k f_k$ will provide an example of a continuous affine function on K for which $\mu(g) > \sup_{x\in K} g(x)$, a fact which contradicts Corollary 4.1.7. $\square$

Two Steffensen–Popoviciu measures μ and ν on K are said to be *equivalent* (abbreviated, $\mu \sim \nu$) provided that

$$\int_K f(x)\, d\mu(x) = \int_K f(x)\, d\nu(x) \quad \text{for all } f \in A(K).$$

Using the density of $E'|_K + \mathbb{R}\cdot 1$ into $A(K)$, we can rewrite the fact that x is the barycenter of μ as

$$\mu \sim \delta_x.$$

We end this section with a monotonicity property.

Proposition 4.1.9 *Suppose that K is as above, μ is a Borel probability measure on K and $f\colon K \to \mathbb{R}$ is a continuous convex function. Then the function*

$$M(t) = \int_K f(tx + (1-t)x_\mu)\, d\mu(x)$$

is convex and nondecreasing on $[0,1]$.

When $E = \mathbb{R}^n$ and μ is the Lebesgue measure, the value of M at t equals the arithmetic mean of $f|_{K_t}$, where K_t denotes the image of K through the mapping $x \to tx + (1-t)x_\mu$, that is,

$$M(t) = \frac{1}{\mu(K_t)} \int_{K_t} f(x)\, d\mu(x).$$

Proposition 4.1.9 tells us that the arithmetic mean of $f|_{K_t}$ decreases to $f(x_\mu)$ when K_t shrinks to x_μ. The proof is based on the following approximation argument:

Lemma 4.1.10 *Every Borel probability measure μ on K is the pointwise limit of a net of discrete Borel probability measures μ_α, each having the same barycenter as μ.*

Proof. We have to prove that for each $\varepsilon > 0$ and each finite family $f_1, \dots, f_n$ of continuous real functions on K there exists a discrete Borel probability measure ν such that

$$x_\nu = x_\mu \quad \text{and} \quad \sup_{1\le k\le n} |\nu(f_k) - \mu(f_k)| < \varepsilon.$$

As K is compact and convex and the functions f_k are continuous, there exists a finite covering $(D_\alpha)_\alpha$ of K by open convex sets such that the oscillation of each of the functions f_k on each set D_α is less than ε. Let $(\varphi_\alpha)_\alpha$ be a partition of unity, subordinated to the covering $(D_\alpha)_\alpha$ and put

$$\nu = \sum_\alpha \mu(\varphi_\alpha)\delta_{x(\alpha)},$$

where $x(\alpha)$ is the barycenter of the measure $f \to \mu(\varphi_\alpha f)/\mu(\varphi_\alpha)$. As D_α is convex and the support of φ_α is included in D_α, we have $x(\alpha) \in \overline{D}_\alpha$. On the other hand,

$$\mu(h) = \sum_\alpha \mu(h\varphi_\alpha) = \sum_\alpha \frac{\mu(h\varphi_\alpha)}{\mu(\varphi_\alpha)}\, \mu(\varphi_\alpha) = \sum_\alpha h(x(\alpha))\, \mu(\varphi_\alpha) = \nu(h)$$

for all continuous affine functions $h\colon K \to \mathbb{R}$. Consequently, μ and ν have the same barycenter. Finally, for each k,

$$\begin{aligned} |\nu(f_k) - \mu(f_k)| &= \Big|\sum_\alpha \mu(\varphi_\alpha) f_k(x(\alpha)) - \sum_\alpha \mu(\varphi_\alpha f_k)\Big| \\ &= \Big|\sum_\alpha \mu(\varphi_\alpha)\Big[f_k(x(\alpha)) - \frac{\mu(\varphi_\alpha f_k)}{\mu(\varphi_\alpha)}\Big]\Big| \\ &\le \varepsilon \cdot \sum_\alpha \mu(\varphi_\alpha) = \varepsilon. \end{aligned}$$

$\square$

Proof of Proposition 4.1.9. A straightforward computation shows that $M(t)$ is convex and $M(t) \leq M(1)$. Then, assuming the inequality $M(0) \leq M(t)$, from the convexity of $M(t)$ we infer

$$\frac{M(t) - M(s)}{t - s} \geq \frac{M(s) - M(0)}{s} \geq 0$$

for all $0 \leq s < t \leq 1$ that is, $M(t)$ is nondecreasing. To end the proof, it remains to show that $M(t) \geq M(0) = f(x_\mu)$. For this, choose a net $(\mu_\alpha)_\alpha$ of discrete Borel probability measures on K, as in Lemma 4.1.10 above. Clearly,

$$f(x_\mu) \leq \int_K f(tx + (1-t)x_\mu)\, d\mu_\alpha(x) \quad \text{for all } \alpha$$

and thus the desired conclusion follows by passing to the limit over α. □

Exercises

1. Prove that $(\prod_{k=1}^n (x_k^2 + a_k))dx_1 \cdots dx_n$ is a Steffensen–Popoviciu measure on $[-1,1]^n$, for all $a_1, \ldots, a_n > -1/3$.

2. Prove that any closed ball in $\mathbb{R}^n$ admits a Steffensen–Popoviciu measure that is not positive.

3. (The failure of Theorem 1.5.7 in higher dimensions) Consider the piecewise linear convex function

$$f(x,y) = \sup\{|x|, |y|, 2|x+y| - 3, 2|x-y| - 3\},$$

defined on the square $|x| \leq 2$, $|y| \leq 2$. Prove that f cannot be represented as a sum (with nonnegative coefficients) of a linear function and functions of the form $|g - \alpha|$, with g linear and $\alpha \in \mathbb{R}$.

4.2 The Jensen–Steffensen Inequality and Majorization

As in the preceding section, K will denote a compact convex subset of a locally convex Hausdorff space E. The following result extends the left-hand side of the Hermite–Hadamard inequality:

Theorem 4.2.1 (The generalized Jensen–Steffensen inequality) *Suppose that μ is a signed Borel measure on K with $\mu(K) > 0$. Then the following assertions are equivalent*:

(i) *μ is a Steffensen–Popoviciu measure*;

(ii) *μ admits a barycenter x_μ and*

$$f(x_\mu) \le \frac{1}{\mu(K)} \int_K f(x)\, d\mu(x).$$

for all continuous convex functions $f\colon K \to \mathbb{R}$.

In order to prove this result we shall need the fact that each continuous convex function on K coincides with its lower envelope.

Given a function f in $C(K)$, we can attach to it a *lower envelope*,

$$\underline{f}(x) = \sup\{h(x) \mid h \in A(K) \text{ and } f \ge h\},$$

and an *upper envelope*,

$$\overline{f}(x) = \inf\{h(x) \mid h \in A(K) \text{ and } h \ge f\}.$$

They are related by formulæ of the form

$$\underline{f} = -\overline{(-f)},$$

so it suffices to investigate the properties of one type of envelope, say the upper one:

Lemma 4.2.2 *The upper envelope $\overline{f}$ is concave, bounded and upper semicontinuous. Moreover*:

(i) $f \le \overline{f}$ *and* $f = \overline{f}$ *if* f *is concave*;

(ii) *if* $f, g \in C(K)$, *then* $\overline{f+g} \le \overline{f} + \overline{g}$ *with equality if* $g \in A(K)$; *also,* $\overline{\alpha f} = \alpha \overline{f}$ *if* $\alpha \ge 0$;

(iii) *the map* $f \to \overline{f}$ *is nonexpansive, that is,* $|\overline{f} - \overline{g}| \le \|f - g\|$.

Proof. Most of this lemma follows directly from the definitions. We shall concentrate here on the less obvious assertion, namely the second part of (i). It may be proved by *reductio ad absurdum.* Assume that $f(x_0) < \overline{f}(x_0)$ for some $x_0 \in K$. By Theorem A.2.4, there exists a closed hyperplane which strictly separates the convex sets $K_1 = \{(x_0, \overline{f}(x_0))\}$ and $K_2 = \{(x, r) \mid f(x) \ge r\}$. This means the existence of a continuous linear functional L on $E \times \mathbb{R}$ and of a scalar λ such that

$$\sup_{(x,r)\in K_2} L(x, r) < \lambda < L(x_0, \overline{f}(x_0)). \tag{4.9}$$

Then $L(x_0, \overline{f}(x_0)) > L(x_0, f(x_0))$, which yields $L(0, 1) > 0$. The function

$$h = \frac{\lambda - L(x, 0)}{L(0, 1)}$$

belongs to $A(K)$ and $L(x, h(x)) = \lambda$ for all x. By (4.9), we infer that $h > f$ and $h(x_0) < \overline{f}(x_0)$, a contradiction. □

Proof of Theorem 4.2.1. The implication (i) $\Rightarrow$ (ii) follows from Lemmas 4.1.6 and 4.2.2. In fact,

$$\begin{aligned} f(x_\mu) &= \sup\{h(x_\mu) \mid h \in A(K), h \leq f\} \\ &= \sup\left\{ \frac{1}{\mu(K)} \int_K h \, d\mu \;\middle|\; h \in A(K), h \leq f \right\} \\ &\leq \frac{1}{\mu(K)} \int_K f \, d\mu. \end{aligned}$$

The implication (ii) $\Rightarrow$ (i) is clear. $\square$

The classical Jensen–Steffensen inequality (see Theorem 1.5.6) represents the case where

$$\mu = \sum_{k=1}^{n} p_k \delta_{x_k}$$

is a discrete measure associated to a family of points $x_1 \leq \cdots \leq x_n$ in an interval $[a, b]$, and to a family of real weights $p_1, \ldots, p_n$ which verify the condition (4.5) above. In fact, this is a Steffensen–Popoviciu measure with barycenter

$$x_\mu = \frac{\sum_{k=1}^{n} p_k x_k}{\sum_{k=1}^{n} p_k},$$

and Theorem 4.2.1 applies.

It is worth noticing that the Jensen–Steffensen inequality also holds under the more general condition (4.4).

The discussion above leaves open the case of signed Borel measures of zero total mass. This is settled by the following result:

Proposition 4.2.3 *If μ is a signed Borel measure on K such that $\mu(K) = 0$ and*

$$\int_K f(x) \, d\mu(x) \geq 0 \quad \textit{for all } f \in \mathrm{Conv}(K), f \geq 0,$$

then

$$\int_K f(x) \, d\mu(x) \geq 0 \quad \textit{for all } f \in \mathrm{Conv}(K).$$

Proof. In fact, by replacing μ by $\mu_\varepsilon = \mu + \varepsilon\delta_z$ (where z is any point of K and $\varepsilon > 0$) we obtain a Steffensen–Popoviciu measure. By Theorem 4.2.1,

$$f(x_{\mu_\varepsilon}) \cdot (\mu(K) + \varepsilon) \leq \int_K f(x) \, d\mu(x) + \varepsilon f(z)$$

for all continuous convex functions f on K, and the conclusion follows by letting $\varepsilon \to 0$. $\square$

Proposition 4.2.3 leads naturally to the extension of the concept of majorization from strings of real numbers to Steffensen–Popoviciu measures. The idea is to see any such string as the support of a discrete probability measure.

Definition 4.2.4 Given two Steffensen–Popoviciu measures μ and ν on K, we say that μ is *majorized* by ν (abbreviated, $\mu \prec \nu$ or $\nu \succ \mu$) if

$$\frac{1}{\mu(K)} \int_K f(x)\, d\mu(x) \leq \frac{1}{\nu(K)} \int_K f(x)\, d\nu(x)$$

for all continuous convex functions $f\colon K \to \mathbb{R}$.

Clearly, we may restrict ourselves to the case of *normalized* measures (that is, of unit total mass). By Proposition 4.2.3, if μ and ν are two signed Borel measures on K with $\mu(K) = \nu(K) = 1$, then the relation of majorization $\mu \prec \nu$ can be derived from the following condition:

$$\int_K f(x)\, d\mu(x) \leq \int_K f(x)\, d\nu(x) \quad \text{for all } f \in \mathrm{Conv}(K), f \geq 0. \tag{4.10}$$

This remark can be converted into more suitable criteria of majorization. For example, when combined with the argument of Lemma 4.1.3, it yields the following result:

Lemma 4.2.5 *Let μ and ν be two normalized signed Borel measures on $[a, b]$ such that*

$$\int_a^t (t-x)\, d\mu(x) \leq \int_a^t (t-x)\, d\nu(x) \quad \textit{and} \quad \int_t^b (x-t)\, d\mu(x) \leq \int_t^b (x-t)\, d\nu(x)$$

for all $t \in [a, b]$. Then $\mu \prec \nu$.

Corollary 4.2.6 (The weighted case of the Hardy–Littlewood–Pólya inequality) *Suppose there are given points $x_1, \ldots, x_n, y_1, \ldots, y_n$ in an interval $[a, b]$, and positive weights $p_1, \ldots, p_n$ such that:*

(i) $x_1 \geq \cdots \geq x_n$;

(ii) $\sum_{k=1}^r p_k x_k \leq \sum_{k=1}^r p_k y_k$ *for all $r = 1, \ldots, n-1$;*

(iii) $\sum_{k=1}^n p_k x_k = \sum_{k=1}^n p_k y_k$.

Then $\sum_{k=1}^n p_k \delta_{x_k} \prec \sum_{k=1}^n p_k \delta_{y_k}$.

Proof. We have to show that $\mu = \sum_{k=1}^n p_k \delta_{x_k}$ and $\nu = \sum_{k=1}^n p_k \delta_{y_k}$ verify the hypotheses of Lemma 4.2.5. For example, if $t \in [a, b]$ and r is the first index such that $t \geq x_r$, then

$$\begin{aligned}
\int_a^t (t-x)\, d\nu(x) - \int_a^t (t-x)\, d\mu(x) &= \sum_{k=1}^n p_k (t - y_k)^+ - \sum_{k=1}^n p_k (t - x_k)^+ \\
&\geq \sum_{k=r}^n p_k (t - y_k) - \sum_{k=r}^n p_k (t - x_k) \\
&= \sum_{k=r}^n p_k (x_k - y_k) \geq 0.
\end{aligned}$$

The result now follows. □

More general results can be proved in the framework of Stieltjes measures.

Theorem 4.2.7 *Let $F, G\colon [a,b] \to \mathbb{R}$ be two functions with bounded variation such that $F(a) = G(a)$. Then, in order that*

$$\int_a^b f(x)\, dF(x) \le \int_a^b f(x)\, dG(x)$$

for all continuous convex functions $f\colon [a,b] \to \mathbb{R}$, it is necessary and sufficient that F and G verify the following three conditions:

$$F(b) = G(b)$$
$$\int_a^x F(t)\, dt \le \int_a^x G(t)\, dt \quad \textit{for all } x \in (a,b)$$
$$\int_a^b F(t)\, dt = \int_a^b G(t)\, dt.$$

Corollary 4.2.8 *Let $f, g \in L^1[a,b]$ be two functions. Then $f dx \prec g dx$ if and only if the following conditions are fulfilled:*

$$\int_a^b f(x)\, dx = \int_a^b g(x)\, dx; \quad \int_a^b x f(x)\, dx = \int_a^b x g(x)\, dx;$$
$$\int_a^x (x-t) f(t)\, dt \le \int_a^x (x-t) g(t)\, dt, \quad \textit{for all } x \in [a,b].$$

The proof of Theorem 4.2.7 is an immediate consequence of the following result due to V. I. Levin and S. B. Stečkin [141]:

Lemma 4.2.9 *Let $F\colon [a,b] \to \mathbb{R}$ be a function with bounded variation such that $F(a) = 0$. Then*

$$\int_a^b f(x)\, dF(x) \ge 0 \quad \textit{for all } f \in \operatorname{Conv}(K)$$

if and only if the following three conditions are fulfilled:

$$F(b) = 0, \quad \int_a^b F(x)\, dx = 0, \quad \textit{and} \quad \int_a^x F(t)\, dt \ge 0 \quad \textit{for all } x \in (a,b).$$

Proof. Via an approximation argument we may restrict to the case where f is also piecewise linear. Then, by using twice the integration by parts, we get

$$\int_a^b f(x)\, dF(x) = -\int_a^b F(x) f'(x)\, dx = \int_a^b \left(\int_a^x F(t)\, dt \right) f''(x)\, dx,$$

whence the sufficiency part. For the necessity, notice that $\int_a^x F(t)\,dt < 0$ for some $x \in (a,b)$ yields an interval I around x on which the integral is still negative. Choosing f such that $f'' = 0$ outside I, the above equalities lead to a contradiction. The necessity of the other two conditions follows by checking our statement for $f = 1,\ -1,\ x-a,\ a-x$ (in this order). $\square$

Finally, let us note that the relation of majorization $\prec$ is a partial ordering on the set of Steffensen–Popoviciu measures on K; use the denseness of the space $\mathrm{Conv}(K) - \mathrm{Conv}(K)$ into $C(K)$. Moreover, according to Theorem 4.2.1,

$$\mu \sim \delta_x \text{ implies } \delta_x \prec \mu.$$

Exercises

1. (G. Szegö) If $a_1 \geq a_2 \geq \cdots \geq a_{2m-1} > 0$ and f is a convex function in $[0, a_1]$, prove that

$$\sum_{k=1}^{2m-1} (-1)^{k-1} f(a_k) \geq f\Big(\sum_{k=1}^{2m-1} (-1)^{k-1} a_k\Big).$$

[*Hint*: Consider the measure $\mu = \sum_{k=1}^{2m-1} (-1)^{k-1} \delta_{a_k}$, whose barycenter is $x_\mu = \sum_{k=1}^{2m-1} (-1)^{k-1} a_k$.]

2. (R. Bellman) Let $a_1 \geq a_2 \geq \cdots \geq a_n \geq 0$ and let f be a convex function on $[0, a_1]$ with $f(0) \leq 0$. Prove that

$$\sum_{k=1}^{n} (-1)^{k-1} f(a_k) \geq f\Big(\sum_{k=1}^{n} (-1)^{k-1} a_k\Big).$$

3. Suppose that $x_1 \leq \cdots \leq x_n$ is a family of points in an interval $[a,b]$, and $p_1, \ldots, p_n$ is a family of real weights. Prove that

$$\sum_{k=1}^{n} p_k f(x_k) \geq 0 \quad \text{for all } f \in \mathrm{Conv}([a,b])$$

if and only if $\sum_{k=1}^n p_k = 0$, $\sum_{k=1}^n p_k x_k = 0$, and

$$\sum_{k=1}^{r} p_k (x_{r+1} - x_k) \geq 0 \quad \text{for all } r \in \{1, \ldots, n-1\}.$$

4. Let $p\colon [0,1] \to \mathbb{R}$ be a continuous function which is nondecreasing on $[0, 1/2]$ and satisfies the condition $f(x) = f(1-x)$. Prove that

$$\int_0^1 f(x) p(x)\,dx \leq \Big(\int_0^1 f(x)\,dx\Big)\Big(\int_0^1 p(x)\,dx\Big)$$

for all $f \in \mathrm{Conv}([0,1])$. Infer that

$$\int_0^1 x(1-x)f(x)\,dx \le \frac{1}{6}\int_0^1 f(x)\,dx$$

and

$$\int_0^\pi f(x)\sin x\,dx \le \frac{2}{\pi}\int_0^\pi f(x)\,dx,$$

provided that f is convex on appropriate intervals.
[*Hint*: It suffices to verify the conditions of Corollary 4.2.8 for $f = p$ and $g = \int_0^1 p(x)\,dx$. The third condition in Corollary 4.2.8 reads as

$$\frac{x^2}{2}\int_0^1 p(t)\,dt \ge \int_0^x \int_0^t p(s)\,dsdt \quad \text{for all } x \in [0,1].$$

For $x \in [0,1/2]$ we have to observe that $\int_0^x p(t)\,dt$ is a convex function on $[0,1/2]$, which yields

$$\frac{1}{x}\int_0^x p(t)\,dt \le 2\int_0^{1/2} p(t)\,dt = \int_0^1 p(t)\,dt \quad \text{for all } x \in [0,1/2]. \quad]$$

4.3 Steffensen's Inequalities

The aim of this section is to prove some inequalities associated to a class of Steffensen–Popoviciu measures which satisfy condition (4.7) above.

Theorem 4.3.1 (Steffensen's inequalities) *Let $g\colon [a,b] \to \mathbb{R}$ be an integrable function such that $\lambda = \int_a^b g(t)\,dt \in (0, b-a]$. Then the following two conditions are equivalent*:

(i) $0 \le \int_a^x g(t)\,dt \le x-a$ *and* $0 \le \int_x^b g(t)\,dt \le b-x$, *for all* $x \in [a,b]$;

(ii) $\int_a^{a+\lambda} f(t)\,dt \le \int_a^b f(t)g(t)\,dt \le \int_{b-\lambda}^b f(t)\,dt$, *for all nondecreasing functions* $f\colon [a,b] \to \mathbb{R}$.

Proof. (i) $\Rightarrow$ (ii) In fact,

$$\int_a^b f(t)g(t)\,dt - \int_a^{a+\lambda} f(t)\,dt = \int_a^{a+\lambda} f(t)(g(t)-1)\,dt + \int_{a+\lambda}^b f(t)g(t)\,dt$$

and the right-hand side equals

$$\begin{aligned}
&\int_a^{a+\lambda} f(t)\,d\Big(\int_a^t g(s)\,ds - t + a\Big) - \int_{a+\lambda}^b f(t)\,d\Big(\int_t^b g(s)\,ds\Big)\\
&\qquad = -\int_a^{a+\lambda}\Big(\int_a^t g(s)\,ds - t + a\Big)\,df(t) + \int_{a+\lambda}^b\Big(\int_t^b g(s)\,ds\Big)\,df(t),
\end{aligned}$$

which gives us the left-hand side inequality of (ii). The other inequality can be obtained in a similar manner.

(ii) $\Rightarrow$ (i) Consider the case of nondecreasing functions $-\chi_{[a,x]}$ and $\chi_{[x,b]}$. $\square$

As was observed in Section 1.5, if $F\colon [a,b] \to \mathbb{R}$ is a convex function (which admits finite derivatives at the endpoints), then

$$\lambda F'(a) \le F(a+\lambda) - F(a) \quad \text{and} \quad F(b) - F(b-\lambda) \le \lambda F'(b)$$

for all $\lambda \in [0, b-a]$. Steffensen's inequalities complement these inequalities as follows: $F(a+\lambda) - F(a)$ is less than or equal to

$$\inf\left\{ \int_a^b F'(t)g(t)\,dt \;\middle|\; g \in L^1[a,b], 0 \le g \le 1, \int_a^b g(t)\,dt = \lambda \right\}$$

and $F(b) - F(b-\lambda)$ is greater than or equal to

$$\sup\left\{ \int_a^b F'(t)g(t)\,dt \;\middle|\; g \in L^1[a,b], 0 \le g \le 1, \int_a^b g(t)\,dt = \lambda \right\}.$$

From Steffensen's inequalities we can derive a stronger version of Iyengar's inequality:

Theorem 4.3.2 *Consider a Riemann integrable function $f\colon [a,b] \to \mathbb{R}$ such that the slopes of the lines AC and CB, joining the endpoints $A(a, f(a))$ and $B(b, f(b))$ of the graph of f to the other points $C(x, f(x))$ of the graph, vary between $-M$ and M. Then:*

$$\left| \frac{1}{b-a} \int_a^b f(x)\,dx - \frac{f(a)+f(b)}{2} \right| \le \frac{M}{4}(b-a) - \frac{(f(b)-f(a))^2}{4M(b-a)}.$$

Proof. According to the trapezoidal approximation, it suffices to consider the case where f is piecewise linear. In that case f is absolutely continuous and it satisfies the inequalities

$$0 \le \int_a^x \frac{f'(t)+M}{2M}\,dt = \frac{f(x)-f(a)+M(x-a)}{2M} \le x-a$$

and

$$0 \le \int_x^b \frac{f'(t)+M}{2M}\,dt = \frac{f(b)-f(x)+M(b-x)}{2M} \le b-x$$

for all $x \in [a,b]$. The proof ends by applying Theorem 4.3.1 to the function $(f'+M)/(2M)$. $\square$

Iyengar's inequality has applications to numerical integration.

Exercises

1. (R. Apéry) Let f be a decreasing function on $(0,\infty)$ and g be a real-valued measurable function on $[0,\infty)$ such that $0 \leq g \leq A$ for a suitable positive constant A. Prove that

$$\int_0^\infty f(x)g(x)\,dx \leq A\int_0^\lambda f(x)\,dx,$$

where $\lambda = (\int_0^\infty g\,dx)/A$.

2. (An extension of Steffensen's inequalities due to J. Pečarić) Let G be an increasing and differentiable function on $[a,b]$ and let $f\colon I \to \mathbb{R}$ be a nonincreasing function, where I is an interval that contains the points a, b, $G(a)$ and $G(b)$. If $G(x) \geq x$ for all x, prove that

$$\int_a^b f(x)G'(x)\,dx \geq \int_{G(a)}^{G(b)} f(x)\,dx.$$

If $G(x) \leq x$ for all x, then the reverse inequality holds. Infer from this result Steffensen's inequalities.

3. Infer from the preceding exercise the following inequality due to C. F. Gauss: Let f be a nonincreasing function on $(0,\infty)$. Then for every $\lambda > 0$,

$$\lambda^2\int_\lambda^\infty f(x)\,dx \leq \frac{4}{9}\int_0^\infty x^2 f(x)\,dx.$$

[*Hint*: Take $G(x) = 4x^3/27\lambda^2 + \lambda$, for $\lambda > 0$.]

4.4 Choquet's Theorem

The aim of this section is to present a full extension of the Hermite–Hadamard inequality (1.18) to the framework of continuous convex functions defined on arbitrary compact convex spaces (when the mean values are computed via Borel measures). We start with the metrizable case, following the classical approach initiated by G. Choquet [53].

Theorem 4.4.1 (Choquet's theorem: The Hermite–Hadamard inequality in the metrizable case) *Let μ be a Borel measure on a metrizable compact convex subset K of a locally convex Hausdorff space E. Then there exists a Borel probability measure λ on K such that the following two conditions are verified:*

(i) $\lambda \succ \mu$ *and λ and μ have the same barycenter;*

(ii) *the set* $\operatorname{Ext} K$, *of all extreme points of K, is a Borel set and λ is concentrated on* $\operatorname{Ext} K$ (*that is,* $\lambda(K \setminus \operatorname{Ext} K) = 0$).

Under the hypotheses of Theorem 4.4.1 we get

$$f(x_\mu) \le \frac{1}{\mu(K)} \int_K f(x)\, d\mu(x) \le \int_{\operatorname{Ext} K} f(x)\, d\lambda(x) \tag{4.11}$$

for every continuous convex function $f\colon K \to \mathbb{R}$, a fact which represents a full extension of (1.18) to the case of metrizable compact convex sets. Notice that the right part of (4.11) reflects the maximum principle for convex functions.

In general, the measure λ is not unique, except for the case of simplices; see [200, Section 10].

As was noticed in Theorem 4.2.1, the left-hand side inequality in 4.11 works in the more general framework of Steffensen-Popoviciu measures. This is no longer true for the right-hand side inequality. In fact, $\mu = \delta_{-1} - \delta_0 + \delta_1$ provides an example of Steffensen-Popoviciu measure on $K = [-1, 1]$ which is not majorized by any Steffensen-Popoviciu measure concentrated on $\operatorname{Ext} K$ (that is, by any convex combination of δ_{-1} and δ_1). However, the right-hand side of (4.11) is known to work for certain signed Borel measures. See Exercise 1.

Proof of Theorem 4.4.1. This will be done in four steps.

Step 1. We start by proving that $\operatorname{Ext} K$ is a countable intersection of open sets (and thus it is a Borel set). Here the assumption on metrizability is essential.

Suppose that the topology of K is given by the metric d and for each integer $n \ge 1$ consider the set

$$K_n = \left\{ x \;\middle|\; x = \frac{y+z}{2}, \text{ with } y, z \in K \text{ and } d(y, z) \ge 1/2^n \right\}.$$

Clearly, $\operatorname{Ext} K = K \setminus \bigcup_n K_n$ and an easy compactness argument shows that each K_n is closed. Consequently, $\operatorname{Ext} K = \bigcap_n \complement K_n$ is a countable intersection of open sets.

Step 2. We may choose a maximal Borel probability measure $\lambda \succ \nu$. To show that Zorn's lemma may be applied, consider a chain $\mathcal{C} = (\lambda_\alpha)_\alpha$ in

$$\mathcal{P} = \{\lambda \mid \lambda \succ \nu, \lambda \text{ Borel probability measure on } K\}.$$

As $(\lambda_\alpha)_\alpha$ is contained in the weak-star compact set

$$\{\lambda \mid \lambda \in C(K), \lambda \ge 0, \lambda(1) = 1\},$$

by a compactness argument we may find a subnet $(\lambda_\beta)_\beta$ which converges to a measure $\tilde{\lambda}$ in the weak-star topology. A moment's reflection shows that $\tilde{\lambda}$ is an upper bound for $\mathcal{C}$. Consequently, we may apply Zorn's lemma to choose a maximal Borel probability measure $\lambda \succ \nu$. It remains to prove that λ does the job.

Step 3. Since K is metrizable, it follows that $C(K)$ (and thus $A(K)$) is separable. This is a consequence of Urysohn's lemma in general topology. See e.g. [124]. Every sequence $(h_n)_n$ of affine functions with $\|h_n\| = 1$, which is dense in the unit sphere of $A(K)$, separates the points of K in the sense that for every $x \neq y$ in K there is an h_n such that $h_n(x) \neq h_n(y)$. Consequently, the function

$$\varphi = \sum_{n=1}^{\infty} 2^{-n} h_n^2$$

is continuous and strictly convex, from which it follows that

$$\mathcal{E} = \{x \mid \varphi(x) = \overline{\varphi}(x)\} \subset \operatorname{Ext} K.$$

In fact, if $x = (y+z)/2$, where y and z are distinct points of K, then the strict convexity of φ implies that

$$\varphi(x) < \frac{\varphi(y) + \varphi(z)}{2} \leq \frac{\overline{\varphi}(y) + \overline{\varphi}(z)}{2} \leq \overline{\varphi}(x).$$

Step 4. As a consequence of the maximality of λ, we shall show that

$$\lambda(\varphi) = \lambda(\overline{\varphi}). \tag{4.12}$$

Then $\overline{\varphi} - \varphi \geq 0$ and $\lambda(\overline{\varphi} - \varphi) = 0$, which yields $\lambda\left(\{x \mid \varphi(x) = \overline{\varphi}(x)\}\right) = 0$. Hence λ is concentrated on $\mathcal{E}$.

The proof of (4.12) is based on Lemma 4.2.2. Consider the sublinear functional $q\colon C(K) \to \mathbb{R}$, given by $q(f) = \lambda(\overline{f})$, and the linear functional L defined on $A(K) + \mathbb{R} \cdot \varphi$ by $L(h + \alpha\varphi) = \lambda(h) + \alpha\lambda(\overline{\varphi})$. By Lemma 4.2.2, if $\alpha \geq 0$, then $L(h + \alpha\varphi) = q(h + \alpha\varphi)$, while if $\alpha < 0$, then

$$0 = \overline{\alpha\varphi - \alpha\varphi} \leq \overline{\alpha\varphi} + \overline{(-\alpha\varphi)} = \overline{\alpha\varphi} - \alpha\overline{\varphi},$$

which yields

$$L(h + \alpha\varphi) = \lambda(h + \alpha\overline{\varphi}) \leq \lambda(\overline{h + \alpha\varphi}) = q(h + \alpha\varphi).$$

By the Hahn–Banach extension theorem, there exists a linear extension ω of L to $C(K)$ such that $\omega \leq q$. If $f \leq 0$, then $\overline{f} \leq 0$, so $\omega(f) \leq q(f) = \lambda(\overline{f}) \leq 0$. Therefore $\omega \geq 0$ and the Riesz–Kakutani representation theorem allows us to identify ω with a suitable Borel probability measure on K.

If f is in $\operatorname{Conv}(K)$, then $-f$ is concave and Lemma 4.2.2 yields

$$\omega(-f) \leq q(-f) = \lambda(\overline{-f}) = \lambda(-f)$$

that is, $\lambda \prec \omega$. Or, λ is maximal, which forces $\omega = \lambda$. Consequently,

$$\lambda(\varphi) = \omega(\varphi) = L(\varphi) = \lambda(\overline{\varphi}),$$

which ends the proof. $\square$

As E. Bishop and K. de Leeuw [30] stated, if K is non-metrizable, then $\operatorname{Ext} K$ need not be a Borel set. However, they were able to prove a Choquet-type theorem. By combining their argument (as presented in [200, Section 4]) with Theorem 4.4.1 above, one can prove the following more general result:

Theorem 4.4.2 (The Choquet–Bishop–de Leeuw theorem) *Let μ be a Borel measure on a compact convex subset K of a locally convex Hausdorff space E. Then there exists a Borel probability measure λ on K such that the following two conditions are fulfilled:*

(i) $\lambda \succ \mu$ *and* λ *and* μ *have the same barycenter;*
(ii) λ *vanishes on every Baire subset of* K *which is disjoint from the set of extreme points of* K*.*

Choquet's theory has deep applications to many areas of mathematics such as function algebras, invariant measures and potential theory. R. R. Phelps' book [200] contains a good account on this matter. We shall add here a few words concerning the connection of Theorem 4.4.1 with some old and new inequalities.

When K is the interval $[a,b]$ endowed with the normalized Lebesgue measure $dx/(b-a)$, then x_μ is exactly the midpoint $(a+b)/2$ and $\operatorname{Ext} K = \{a,b\}$. Any Borel probability measure λ concentrated on $\operatorname{Ext} K$ is necessarily a convex combination of Dirac measures,

$$\lambda = \alpha\delta_a + (1-\alpha)\delta_b$$

for some $\alpha \in [0,1]$. Checking the right-hand side inequality in (4.11) for $f = x-a$ and $f = b-x$ we get

$$1-\alpha \geq 1/2 \quad \text{and} \quad \alpha \geq 1/2,$$

that is, $\alpha = 1/2$. Consequently, in this case (4.11) coincides with (1.18) and we conclude that Theorem 4.4.1 provides a full generalization of the Hermite–Hadamard inequality.

In the same way we can infer from Theorem 4.4.1 the following result:

Theorem 4.4.3 *Let f be a continuous convex function defined on an n-dimensional simplex $K = [a_0,\ldots,a_n]$ in $\mathbb{R}^n$ and let μ be a Borel measure on K. Then*

$$\begin{aligned} f(x_\mu) &\leq \frac{1}{\mu(K)} \int_K f(x)\,d\mu \\ &\leq \frac{1}{\operatorname{Vol}_n(K)} \sum_{k=0}^{n} \operatorname{Vol}_n([a_0,\ldots,\widehat{a_k},\ldots,a_n] \cdot f(a_k). \end{aligned}$$

Here $[a_0,\ldots,\widehat{a_k},\ldots,a_n]$ denotes the subsimplex obtained by replacing a_k by x_μ; this is the subsimplex opposite to a_k, when adding x_μ as a new vertex.

Corollary 4.4.4 (The weighted form of the Hermite–Hadamard inequality) *For every continuous convex function $f\colon [a,b] \to \mathbb{R}$ and every Borel measure μ on $[a,b]$, we have*

$$f(x_\mu) \leq \frac{1}{\mu([a,b])} \int_a^b f(x)\,d\mu(x) \leq \frac{b-x_\mu}{b-a} \cdot f(a) + \frac{x_\mu - a}{b-a} \cdot f(b), \tag{4.13}$$

where

$$x_\mu = \frac{1}{\mu([a,b])}\int_a^b x\,d\mu(x)$$

represents the barycenter of μ.

When $d\mu(x) = p(x)\,dx$, for some nonnegative continuous function $p(x)$ whose graph is symmetric with respect to the middle point $(a+b)/2$, then $x_\mu = (a+b)/2$ and Corollary 4.4.4 reduces to a result due to L. Fejér.

In the case of closed balls $K = \overline{B}_R(a)$ in $\mathbb{R}^n$, $\operatorname{Ext} K$ coincides with the sphere $S_R(a)$. According to Theorem 4.4.1, if $f\colon \overline{B}_R(a) \to \mathbb{R}$ is a continuous convex function and μ is the normalized Lebesgue measure on $\overline{B}_R(a)$, then

$$f(a) \le \frac{1}{\operatorname{Vol}\overline{B}_R(a)}\iiint_{\overline{B}_R(a)} f(x)\,dV \le \frac{1}{\operatorname{Area} S_R(a)}\iint_{S_R(a)} f(x)\,dS. \quad (4.14)$$

A similar result works in the case of subharmonic functions, see the Comments at the end of this chapter. As noticed by P. Montel [171], in the context of C^2-functions on open convex sets in $\mathbb{R}^n$, the class of subharmonic functions is strictly larger than the class of convex function. For example, the function $2x^2 - y^2$ is subharmonic but not convex on $\mathbb{R}^2$.

Many interesting inequalities relating weighted means represent averages over the $(n-1)$-dimensional simplex:

$$\Delta_n = \{u = (u_1,\dots,u_n) \mid u_1,\dots,u_n \ge 0,\ u_1 + \cdots + u_n = 1\}.$$

Clearly, Δ_n is compact and convex and its extreme points are the "corners" $(1,0,\dots,0),\dots,(0,0,\dots,1)$.

An easy consequence of Theorem 4.4.1 is the following refinement of the Jensen–Steffensen inequality for functions on intervals:

Theorem 4.4.5 *Suppose that* $f\colon [a,b] \to \mathbb{R}$ *is a continuous convex function. Then for every* n*-tuple* $x = (x_1,\dots,x_n)$ *of elements of* $[a,b]$ *and every Borel measure* μ *on* Δ_n *we have*

$$f\Big(\sum_{k=1}^n w_k x_k\Big) \le \frac{1}{\mu(\Delta_n)}\int_{\Delta_n} f(x\cdot u)\,d\mu \le \sum_{k=1}^n w_k f(x_k). \quad (4.15)$$

Here $(w_1,\dots,w_n)$ *denotes the barycenter of* Δ_n *with respect to* μ. *The above inequalities should be reversed if* f *is concave on* $[a,b]$.

Under the hypotheses of Theorem 4.4.5, the *weighted identric mean* $I(x,\mu)$ is defined by the formula

$$I(x,\mu) = \exp\int_{\Delta_n} \ln(x\cdot u)\,d\mu(u)$$

and the *weighted logarithmic mean* $L(x,\mu)$ is defined by the formula

$$L(x,\mu)=\left(\int_{\Delta_n}\frac{1}{x\cdot u}\,d\mu(u)\right)^{-1}.$$

By (4.15), we infer easily that $L(x,\mu)\le I(x,\mu)$ and that both lie between the *weighted arithmetic mean* $A(x,\mu)=\sum_{k=1}^n w_k x_k$ and the *weighted geometric mean* $G(x,\mu)=\prod_{k=1}^n x_k^{w_k}$, that is,

$$G(x,\mu)\le L(x,\mu)\le I(x,\mu)\le A(x,\mu), \tag{4.16}$$

a fact which constitutes the *weighted geometric-logarithmic-identric-arithmetic mean inequality.*

An important example of a Borel probability measure on Δ_n is the *Dirichlet measure* of parameters $p_1,\dots,p_n>0$,

$$\frac{\Gamma(p_1+\cdots+p_n)}{\Gamma(p_1)\cdots\Gamma(p_n)}x_1^{p_1-1}\cdots x_{n-1}^{p_{n-1}-1}(1-x_1-\cdots-x_{n-1})^{p_n-1}\,dx_1\cdots dx_{n-1}.$$

Its barycenter is the point $(\sum_{k=1}^n p_k)^{-1}\cdot(p_1,\dots,p_n)$.

Exercises

1. (A. M. Fink [81]) Let f be a convex function in $C^2([a,b])$ and let μ be a Borel measure on $[a,b]$ such that $\mu([a,b])>0$ and the solution $y=y(x)$ of the boundary value problem $y''=p$, $y(a)=y(b)=0$, is ≤ 0 on $[a,b]$.

(i) Prove that

$$\frac{1}{\mu([a,b])}\int_a^b f(x)\,d\mu(x)\le\frac{b-x_\mu}{b-a}\cdot f(a)+\frac{x_\mu-a}{b-a}\cdot f(b),$$

where $x_\mu=\int_a^b x\,d\mu(x)/\mu([a,b])$.

(ii) Consider the particular case where $[a,b]=[-1,1]$ and $d\mu(x)=(x^2-1/6)\,dx$. Prove that $y(x)=x^2(x^2-1)/12\le 0$ and $x_\mu=0$, hence

$$\int_{-1}^1 f(x)\left(x^2-\frac{1}{6}\right)dx\le\frac{f(-1)+f(1)}{6}$$

for all convex functions f in $C^2([-1,1])$.

(iii) Consider the particular case where $[a,b]=[-1,1]$ and $d\mu(x)=(x^2-x)\,dx$. Prove that $y=(x^2-1)(x-1)^2/12\le 0$ and $x_\mu=-1$, hence

$$\int_{-1}^1 f(x)(x^2-x)\,dx\le\frac{2}{3}\,f(-1)$$

for all convex functions f in $C^2([-1,1])$. Notice that $(x^2-x)\,dx$ is not a Steffensen–Popoviciu measure on $[-1,1]$.

[*Hint*: Let $G(x,t)$ be the Green's function for the boundary value problem in the statement. Then $G(x,t) = G(t,x)$ and

$$y(x) = \int_a^b G(x,t)\, d\mu(t)$$

so that we can compute $\int_a^b f''(x)y(x)\,dx$ by using the Fubini theorem. To end the proof, notice that

$$\int_a^b G(t,x)f''(x)\,dx = f(t) - f(a)\,\frac{b-t}{b-a} - f(b)\,\frac{t-a}{b-a}. \ ]$$

2. (A higher dimensional analogue of the Hermite–Hadamard inequality) Let f be a continuous concave function defined on a compact convex subset $K \subset \mathbb{R}^n$ of positive volume. Prove that

$$\frac{1}{n+1} \sup_{x\in K} f(x) + \frac{n}{n+1} \inf_{x\in \operatorname{Ext} K} f(x) \le \frac{1}{\operatorname{Vol}_n(K)} \int_K f(x)\,dx \le f(x_k),$$

where x_k is the barycenter of K.

3. (R. R. Phelps [200]) Let E be a normed linear space. Prove that

$$x_n \to x \quad \text{weakly in } E$$

if and only if the sequence $(x_n)_n$ is norm bounded and $\lim_{n\to\infty} f(x_n) = f(x)$ for each extreme point f of the closed unit ball in E'.
[*Hint*: Let K be the closed unit ball in E'. Then K is convex and weak-star compact (see Theorem A.1.6). Each point $x \in E$ gives rise to an affine mapping $A_x\colon K \to \mathbb{R}$, $A_x(x') = x'(x)$. Then apply Theorem 4.4.2.]

4. (R. Haydon) Let E be a real Banach space and let K be a weak-star compact convex subset of E' such that $\operatorname{Ext} K$ is norm separable. Prove that K is the norm closed convex hull of $\operatorname{Ext} K$ (and hence is itself norm separable).

5. Let K be a nonempty compact convex set in a locally convex Hausdorff space E. Given $f \in C(K)$, prove that:

(i) $\overline{f}(x) = \inf\{g(x) \mid g \in -\operatorname{Conv}(K) \text{ and } g \ge f\}$;

(ii) for each pair of functions $g_1, g_2 \in -\operatorname{Conv}(K)$ with $g_1, g_2 \ge f$, there is a function $g \in -\operatorname{Conv}(K)$ such that $g_1, g_2 \ge g \ge f$;

(iii) $\mu(\overline{f}) = \inf\{\mu(g) \mid g \in -\operatorname{Conv}(K) \text{ and } g \ge f\}$.

6. (G. Mokobodzki) Infer from the preceding exercise that a Borel probability measure μ on K is maximal if and only if $\mu(f) = \mu(\overline{f})$ for all continuous convex functions f on K (equivalently, for all functions $f \in C(K)$).

4.5 Comments

The highlights of classical Choquet theory have been presented by R. R. Phelps [200]. However, the connection between the Hermite–Hadamard inequality and Choquet's theory remained unnoticed until very recently. In 2001, during a conference presentation at the University of Timisoara, C. P. Niculescu called the attention to this matter and sketched the theory of Steffensen–Popoviciu measures. Details appeared in [181] and [182]. Unfortunately, the claim of Theorems 4 and 5 in [181] on the existence of Borel probability measures majorizing a given Steffensen-Popoviciu measure is false. This leaves open the extension of Theorem 4.1.1 to the case of signed Borel measures.

Proposition 4.1.9 was first noticed by S. S. Dragomir in a special case. In its present form it is due to C. P. Niculescu [180].

The Steffensen inequalities appeared in his paper [228]. Using the right-hand side inequality in Theorem 4.3.1 (ii), he derived in [229] what is now known as the *Jensen–Steffensen inequality.* The proof of Theorem 4.3.1 which appears in this book is due to P. M. Vasić and J. Pečarić. See [196, Section 6.2].

K. S. K. Iyengar published his inequality in [113]. Its generalization, as presented in Theorem 4.3.2, follows the paper by C. P. Niculescu and F. Popovici [187].

As noticed in Theorem 1.10.1, the relation of majorization $x \prec y$ can be characterized by the existence of a doubly stochastic matrix P such that $x = Py$. Thinking of x and y as discrete probability measures, this fact can be rephrased as saying that y is a dilation of x. The book by R. R. Phelps [200] indicates the details of an extension (due to P. Cartier, J. M. G. Fell and P. A. Meyer) of this characterization to the general framework of Borel probability measures on compact convex sets (in a locally convex Hausdorff space).

Related to the relation of majorization is the notion of Schur convexity. Let D be an open convex subset of $\mathbb{R}^n$ which is symmetric, that is, invariant under each permutation of the coordinates. A function $f\colon D \to \mathbb{R}$ is said to be *Schur convex* (or Schur increasing) if it is nondecreasing relative to $\prec$. Similarly for *Schur concave* functions, also called Schur decreasing. A Schur convex function is always symmetric. An obvious example is

$$F(x_1, \ldots, x_n) = \sum_{k=1}^{n} f(x_k)$$

where $f\colon \mathbb{R} \to \mathbb{R}$ is a convex function. More generally, every symmetric and convex (concave) function on D is Schur convex (concave).

A symmetric C^1-function f on D is Schur convex if and only if $\frac{\partial f}{\partial x_i} - \frac{\partial f}{\partial x_j}$ is always of the same sign as $x_i - x_j$, for all $i, j \in \{1, \ldots, n\}$.

The area of a triangle is a Schur concave function of its sides. The radius of the circumscribed circle of a triangle is a Schur convex function of its sides.

The books by A. W. Marshall and I. Olkin [155] and Y. L. Tong [237] contain significant applications of Schur convexity.

Following our paper [186], we shall show that the Hermite–Hadamard inequality also works in the context of subharmonic functions. The key remark is the possibility of extending the result of Exercise 4 in Section 1.9 to several variables.

Let Ω be a bounded open subset of $\mathbb{R}^n$ with smooth boundary. Then the Dirichlet problem

$$\begin{cases} \Delta\varphi = 1 & \text{on } \Omega \\ \varphi = 0 & \text{on } \partial\Omega \end{cases} \tag{4.17}$$

has a unique solution, which is negative on Ω, according to the maximum principle for elliptic problems. See [211]. By Green's formula, for every u in $C^2(\Omega) \cap C^1(\overline{\Omega})$ we have

$$\int_\Omega \begin{vmatrix} u & \varphi \\ \Delta u & \Delta\varphi \end{vmatrix} dV = \int_{\partial\Omega} \begin{vmatrix} u & \varphi \\ \nabla u & \nabla\varphi \end{vmatrix} \cdot n\, dS,$$

that is, in view of (4.17),

$$\begin{aligned} \int_\Omega u\, dV &= \int_\Omega u\Delta\varphi\, dV \\ &= \int_\Omega \varphi\Delta u\, dV + \int_{\partial\Omega} u(\nabla\varphi\cdot n)\, dS - \int_{\partial\Omega} \varphi(\nabla u\cdot n)\, dS \\ &= \int_\Omega \varphi\Delta u\, dV + \int_{\partial\Omega} u(\nabla\varphi\cdot n)\, dS \end{aligned}$$

for every $u \in C^2(\Omega) \cap C^1(\overline{\Omega})$. We are then led to the following result:

Theorem (The Hermite–Hadamard inequality for subharmonic functions) *If $u \in C^2(\Omega) \cap C^1(\overline{\Omega})$ is subharmonic (that is, $\Delta u \geq 0$ on Ω) and φ satisfies* (4.17), *then*

$$\int_\Omega u\, dV < \int_{\partial\Omega} u(\nabla\varphi\cdot n)\, dS$$

except for harmonic functions (when equality occurs).

The equality case needs the remark that $\int_\Omega \varphi\Delta u\, dV = 0$ yields $\varphi\Delta u = 0$ on Ω, and thus $\Delta u = 0$ on Ω; notice that $\varphi\Delta u$ is continuous and $\varphi\Delta u \leq 0$ since $\varphi < 0$ on Ω.

In the case of balls $\Omega = B_R(a)$ in $\mathbb{R}^3$, the solution of the problem (4.17) is $\varphi(x) = (\|x\|^2 - R^2)/6$ and $\nabla\varphi\cdot n = x/3 \cdot x/\|x\| = \|x\|/3$, so that by combining the maximum principle for elliptic problems with the conclusion of the above theorem we obtain the following Hermite–Hadamard type inequality for subharmonic functions:

$$u(a) \leq \frac{1}{\operatorname{Vol} \overline{B}_R(a)} \iiint_{\overline{B}_R(a)} u(x)\, dV < \frac{1}{\operatorname{Area} S_R(a)} \iint_{S_R(a)} u(x)\, dS$$

for every $u \in C^2(B_R(a)) \cap C^1(\overline{B}_R(a))$ with $\Delta u \geq 0$, which is not harmonic.

Consider now the Green kernel $G(x,y)$ associated with $-\Delta$ on Ω. The solution $u \in C(\overline{\Omega}) \cap C^2(\Omega)$ of the Dirichlet problem

$$\begin{cases} \Delta u = f & \text{on } \Omega \\ u = 0 & \text{on } \partial\Omega \end{cases} \tag{4.18}$$

where $f \in L^1(\Omega)$, and $f \geq 0$, can be represented as

$$u(x) = \int_\Omega G(x,y) f(y)\, dV. \tag{4.19}$$

By varying f, the set of all such functions u constitutes a subcone $\mathcal{SH}_0^+(\Omega)$, of the convex cone $\mathcal{SH}^+(\Omega)$ of all nonnegative superharmonic functions on Ω. Recall that a function u is called *superharmonic* if $-u$ is subharmonic. The maximum principle for elliptic problems assures that $u \geq 0$ (and the same is true for G).

Theorem (The extension of Berwald's inequality for subharmonic functions) *Assume that* $0 < r \leq 1 \leq s$ *and*

$$C = C(r,s;\mu,\nu) = \sup_{y \in \Omega} \left(\int_\Omega G(x,y)^s d\mu(x) \right)^{1/s} \Big/ \left(\int_\Omega G(x,y)^r d\nu(x) \right)^{1/r} < \infty,$$

where μ *and* ν *are two Borel probability measures on* Ω. *Then*

$$\left(\int_\Omega u^s(x) d\mu(x) \right)^{1/s} \leq C \left(\int_\Omega u^r(x) d\nu(x) \right)^{1/r} \tag{4.20}$$

for every $u \in \mathcal{SH}_0^+(\Omega)$ *and the constant* $C = C(r,s;\mu,\nu)$ *is sharp.*

If μ and ν are absolutely continuous with respect to the Lebesgue measure on Ω, then the inequality (4.20) extends (by density) to the whole cone $\mathcal{SH}^+(\Omega)$.

Proof. We use the representation formula (4.19). Then, by applying the Rogers-Hölder inequality, the Fubini theorem and finally the Minkowski inequality, we get

$$\begin{aligned}
\int_\Omega u^s(x)d\mu(x) &= \int_\Omega u^{s-1}(x)\left(\int_\Omega G(x,y)f(y)dV\right)d\mu(x)\\
&= \int_\Omega\left(\int_\Omega G(x,y)u^{s-1}(x)d\mu(x)\right)f(y)\,dV\\
&\le \int_\Omega\left(\int_\Omega G(x,y)^s d\mu(x)\right)^{1/s}\cdot\left(\int_\Omega u^{(s-1)s'}(x)d\mu(x)\right)^{1/s'} f(y)\,dV\\
&\le C\left(\int_\Omega u^s(x)d\mu(x)\right)^{1/s'}\int_\Omega\left(\int_\Omega G(x,y)^r d\nu(x)\right)^{1/r} f(y)\,dV\\
&\le C\left(\int_\Omega u^s(x)d\mu(x)\right)^{1/s'}\left(\int_\Omega\left(\int_\Omega G(x,y)f(y)dV\right)^r d\nu(x)\right)^{1/r}\\
&\le C\left(\int_\Omega u^s(x)d\mu(x)\right)^{1/s'}\left(\int_\Omega u^r(x)d\nu(x)\right)^{1/r},
\end{aligned}$$

where s' is the conjugate of s. This yields the formula (4.20). The fact that $C = C(r,s;\mu,\nu)$ is sharp follows by considering the case of functions $u\,(x) = G(x,y)$, for $y \in \Omega$ arbitrarily fixed. $\square$

The result of theorem above is valid for every function u representable via nonnegative kernels by formulae of the type (4.19), with f continuous and nonnegative.

For $\Omega = (a,b)$, the Green kernel is

$$G(x,y) = \begin{cases} (y-a)(b-x) & \text{if } a \le y \le x \le b \\ (x-a)(b-y) & \text{if } a \le x \le y \le b \end{cases}$$

and thus for $d\mu(x) = d\nu(x) = dx/(b-a)$ we have

$$C(r,s;dx/(b-a),dx/(b-a)) = (r+1)^{1/r}\,/\,(s+1)^{1/s}\,.$$

This allows us to recover Berwald's inequality in the range $0 < r \le 1 \le s < \infty$, for continuous concave functions of a real variable.

As noticed by G. Choquet [53], the similarities between the convex functions and the subharmonic functions can be explained by the existence of a much larger theory concerning the pairs $(X,\mathcal{S})$, where X is a Hausdorff compact space and $\mathcal{S}$ is a convex cone of lower semicontinuous and bounded below functions $f\colon X \to \mathbb{R} \cup \{\infty\}$. Within this framework, the role of the relation of majorization on the set $\operatorname{Prob}(X)$, of all Borel probability measures on X, is played by the relation

$$\mu \le_{\mathcal{S}} \nu \quad \text{if and only if } \mu(s) \le \nu(s) \quad \text{for all } s \in \mathcal{S}.$$

A

Background on Convex Sets

The Hahn–Banach theorem is a deep result in functional analysis which provides important consequences to convex function theory. Its proof and some of its applications are presented here for the convenience of the reader.

A.1 The Hahn–Banach Extension Theorem

Throughout, E will denote a real linear space.

A functional $p\colon E \to \mathbb{R}$ is *subadditive* if $p(x+y) \leq p(x) + p(y)$ for all $x, y \in E$; p is *positively homogeneous* if $p(\lambda x) = \lambda p(x)$ for each $\lambda \geq 0$ and each x in E; p is *sublinear* if it has both the above properties. A sublinear functional p is a seminorm if $p(\lambda x) = |\lambda| p(x)$ for all scalars. Finally, a seminorm p is a norm if

$$p(x) = 0 \implies x = 0.$$

If p is a sublinear functional, then $p(0) = 0$ and $-p(-x) \leq p(x)$. If p is a seminorm, then $p(x) \geq 0$ for all x in E and $\{x \mid p(x) = 0\}$ is a linear subspace of E.

Theorem A.1.1 (The Hahn–Banach theorem) *Let p be a sublinear functional on E, let E_0 be a linear subspace of E, and let $f_0\colon E_0 \to \mathbb{R}$ be a linear functional dominated by p, that is, $f_0(x) \leq p(x)$ for all $x \in E_0$. Then f_0 has a linear extension f to E which is also dominated by p.*

Proof. We consider the set $\mathcal{P}$ of all pairs (h, H), where H is a linear subspace of E that contains E_0 and $h\colon H \to \mathbb{R}$ is a linear functional dominated by p that extends f_0. $\mathcal{P}$ is nonempty (as $(f_0, E_0) \in \mathcal{P}$). One can easily prove that $\mathcal{P}$ is inductively ordered with respect to the order relation

$$(h, H) \prec (h', H') \iff H \subset H' \text{ and } h'|_H = h,$$

so that by Zorn's lemma we infer that $\mathcal{P}$ contains a maximal element (g, G). It remains to prove that $G = E$.

If $G \neq E$, then we can choose an element $z \in E \backslash G$ and denote by G' the set of all elements of the form $x + \lambda z$, with $x \in G$ and $\lambda \in \mathbb{R}$. Clearly, G' is a linear space that contains G strictly and the formula

$$g'(x + \lambda z) = g(x) + \alpha\lambda$$

defines (for every $\alpha \in \mathbb{R}$) a linear functional on G' that extends g. We shall show that α can be chosen so that g' is dominated by p (a fact that contradicts the maximality of (g, G)).

In fact, g' is dominated by p if

$$g(x) + \alpha\lambda \leq p(x + \lambda z)$$

for every $x \in G$ and every $\lambda \in \mathbb{R}$. If $\lambda \geq 0$, this means:

$$g(x) + \alpha \leq p(x + z) \quad \text{for every } x \in G.$$

If $\lambda < 0$, we get (after simplification by $-\lambda$),

$$g(x) - \alpha \leq p(x - z) \quad \text{for every } x \in G.$$

Therefore, we have to choose α such that

$$g(u) - p(u - z) \leq \alpha \leq p(v + z) - g(v)$$

for every $u, v \in G$. This choice is possible because

$$g(u) + g(v) = g(u + v) \leq p(u + v) \leq p(u - z) + p(v + z)$$

for all $u, v \in G$, which yields

$$\sup_{u \in G}(g(u) - p(u - z)) \leq \inf_{v \in G}(p(v + z) - g(v)).$$

The proof is now complete. $\square$

Corollary A.1.2 *If p is a sublinear functional on a real linear space E, then for every element $x_0 \in E$ there exists a linear functional $f\colon E \to \mathbb{R}$ such that $f(x_0) = p(x_0)$ and $f(x) \leq p(x)$ for all x in E.*

Proof. Take $E_0 = \{\lambda x_0 \mid \lambda \in \mathbb{R}\}$ and $f_0(\lambda x_0) = \lambda p(x_0)$ in Theorem A.1.1. $\square$

The continuity of a linear functional on a topological linear space E means that it is bounded in a neighborhood of the origin. We shall denote by E' the *dual space* of E that is, the space of all continuous linear functionals on E.

In the context of normed linear spaces, the remark above allows us to define the norm of a continuous linear functional $f\colon E \to \mathbb{R}$ by the formula

$$\|f\| = \sup_{\|x\|\leq 1} |f(x)|.$$

With respect to this norm, the dual space of a normed linear space is always complete.

It is worth noting the following variant of Theorem A.1.1 in the context of real normed linear spaces:

Theorem A.1.3 (The Hahn–Banach theorem) *Let E_0 be a linear subspace of the normed linear space E, and let $f_0 \colon E_0 \to \mathbb{R}$ be a continuous linear functional. Then f_0 has a continuous linear extension f to E, with $\|f\| = \|f_0\|$.*

Corollary A.1.4 *If E is a normed linear space, then for each $x_0 \in E$ with $x_0 \neq 0$ there exists a continuous linear functional $f \colon E \to \mathbb{R}$ such that $f(x_0) = \|x_0\|$ and $\|f\| = 1$.*

Corollary A.1.5 *If E is a normed linear space and x is an element of E such that $f(x) = 0$ for all f in the dual space of E, then $x = 0$.*

The *weak topology* on E is the locally convex topology associated to the family of seminorms

$$p_F(x) = \sup\{|f(x)| \bigm| f \in F\},$$

where F runs over all nonempty finite subsets of E'. A sequence $(x_n)_n$ converges to x in the weak topology (abbreviated, $x_n \overset{w}{\to} x$) if and only if $f(x_n) \to f(x)$ for every $f \in E'$. When $E = \mathbb{R}^n$ this is the coordinate-wise convergence and agrees with the norm convergence. In general, the norm function is only *weakly lower semicontinuous*, that is,

$$x_n \overset{w}{\to} x \implies \|x\| \leq \liminf_{n\to\infty} \|x_n\|.$$

By Corollary A.1.5 it follows that E' *separates* E in the sense that

$$x, y \in E \quad \text{and} \quad f(x) = f(y) \quad \text{for all } f \in E' \Longrightarrow x = y.$$

As a consequence we infer that the weak topology is separated (equivalently, Hausdorff).

For E' we can speak of the normed topology, of the weak topology (associated to $E'' = (E')'$) and also of the *weak-star topology*, which is associated to the family of seminorms p_F defined as above, with the difference that F runs over all nonempty finite subsets of E. The weak-star topology on E' is separated.

A net $(f_i)_{i\in I}$ (over some directed set I) converges to f in the weak-star topology (abbreviated, $f_i \overset{w^*}{\to} f$) if and only if $f_i(x) \to f(x)$ for all $x \in E$.

Theorem A.1.6 (The Banach–Alaoglu theorem) *If E is a normed linear space, then the closed unit ball of its dual space is compact in the weak-star topology. Consequently, each net of points of this ball has a converging subnet.*

See [64, p. 47] for details.

When E is a separable normed linear space, the closed unit ball of E' is also a metrizable space in the weak-star topology (and in this case dealing with sequences suffices as well). We come to the separability situation very often, by replacing E with a subspace generated by a suitable sequence of elements.

Remark A.1.7 According to the Banach–Alaoglu theorem, if E is a normed linear space, then each weak-star closed subset of the closed unit ball of the dual of E is weak-star compact. This is a big source of compact convex sets in mathematics. For example, so is the set $\mathrm{Prob}(X)$, of all Borel probability measures on a compact Hausdorff space X. These are the regular σ-additive measures μ on the Borel subsets of X with $\mu(X) = 1$. The Riesz–Kakutani representation theorem (see [103, p. 177]) allows us to identify $\mathrm{Prob}(X)$ with the following weak-star closed subset of norm-1 functionals of $C(X)'$:

$$K = \{L \mid L \in C(X)',\ L(1) = 1 = \|L\|\}.$$

Notice that K consists of *positive functionals*, that is,

$$f \in C(X), \quad f \geq 0 \text{ implies } L(f) \geq 0.$$

In fact, if the range of f is included in $[0, 2r]$, then $\|f - r\| \leq r$, so that $r \geq |L(f - r)| = |L(f) - r|$, that is, $L(f) \in [0, 2r]$.

Corollary A.1.4 yields an important canonical embedding of each normed linear space E into its second dual E'':

$$J_E \colon E \to E'', \quad J_E(x)(x') = x'(x).$$

One can easily show that J_E is a linear isometry.

A Banach space E is said to be *reflexive* if J_E is onto (that is, if E is isometric with its second dual through J_E). Besides the finite dimensional Banach spaces, other examples of reflexive Banach spaces are Hilbert spaces and the spaces $L^p(\mu)$ for $1 < p < \infty$. One can easily prove the following permanence properties:

(R1) every closed subspace of a reflexive space is reflexive;
(R2) the dual of a reflexive space is also a reflexive space;
(R3) reflexivity preserves under renorming by an equivalent norm.

Property (R3) follows from the following characterization of reflexivity:

Theorem A.1.8 (The Eberlein–Šmulyan theorem) *A Banach space E is reflexive if and only if every bounded sequence of elements of E admits a weakly converging subsequence.*

The *necessity* part is a consequence of the Banach–Alaoglu theorem (Theorem A.1.6). In fact, we may restrict ourselves to the case where E is also separable. The *sufficiency* part follows from the remark that J_E maps the closed unit ball of E into a w'-dense (and also w'-closed) subset of the closed unit ball of E''. Full details are available in books such as those by H. W. Alt [6], J. B. Conway [58] or M. M. Day [64].

A.2 Separation of Convex Sets

The notion of a hyperplane in a real linear space E was introduced in Section 3.3 as the translate of the kernel of a nonzero linear functional. It can be equally defined as a maximal proper affine subset. In fact, if $h\colon E \to \mathbb{R}$ is a nonzero linear functional, we may choose a $v \in E$ with $h(v) = 1$. Then all $x \in E$ can be represented as

$$x = (x - h(x)v) + h(x)v$$

where $x - h(x)v \in \ker h$. This shows that $\ker h$ is a linear space of codimension 1, and thus all its translates are maximal proper affine subsets.

Conversely, if H is a maximal proper affine set in E and $x_0 \in H$, then $-x_0 + H$ is a linear subspace (necessarily of codimension 1). Hence there exists a vector $v \neq 0$ such that E is the direct sum of $-x_0 + H$ and $\mathbb{R}v$, that is, all $x \in E$ can be uniquely represented as

$$x = (-x_0 + y) + \lambda v$$

for suitable $y \in H$ and $\lambda \in \mathbb{R}$. The formula $h(x) = \lambda$ defines a linear functional h such that $h(v) = 1$ and $h(x) = 0$ if and only if $x \in -x_0 + H$. Consequently,

$$H = \{x \mid h(x) = h(x_0)\}.$$

Suppose now that E is a Hausdorff linear topological space. Then the discussion above shows that the closed hyperplanes H in E coincide with the constancy sets of nonzero continuous and linear functionals. In fact, it suffices to consider the case where H is a closed subspace of codimension 1. In that case E/H is 1-dimensional and thus it is algebraically and topologically isomorphic to $\mathbb{R}$. By composing such an isomorphism with the canonical projection from E onto E/H we obtain a continuous linear functional h for which $H = \ker h$.

To each hyperplane $\{x \mid h(x) = \lambda\}$ we can attach two half-spaces,

$$\{x \mid h(x) \leq \lambda\} \quad \text{and} \quad \{x \mid h(x) \geq \lambda\}.$$

We say that two sets A and B are *separated* by a hyperplane H if they are contained in different half-spaces. The separation is *strict* if at least one of the two sets does not intersect H.

A basic result concerning the separability by hyperplanes is as follows:

Theorem A.2.1 (Mazur's theorem) *Let K be a convex set with nonempty interior in a real linear topological Hausdorff space E and let A be an affine subset which contains no interior point of K. Then there exists a closed hyperplane H such that $H \supset A$ and $H \cap K = \emptyset$.*

In other words, there exists a continuous linear functional $h\colon E \to \mathbb{R}$ and a number $\alpha \in \mathbb{R}$ such that $h(x) = \alpha$ if $x \in A$ and $h(x) < \alpha$ if $x \in K$.

Proof. We may assume that K is a convex neighborhood of the origin since otherwise we choose an interior point x_0 in K and replace K and A by $K - x_0$ and $A - x_0$ respectively. Notice that translations are isomorphisms, so they preserve the nature of K and A. Denote by E_0 the linear span of A. Then A is a hyperplane in E_0, which yields a linear functional $f_0\colon E_0 \to \mathbb{R}$ such that

$$A = \{x \in E_0 \mid f_0(x) = 1\}.$$

The Minkowski functional of K,

$$p_K(x) = \inf\{\lambda > 0 \mid x \in \lambda K\}$$

is sublinear and $\{x \mid p_k(x) < 1\}$ coincides with the interior of K. In fact, if x is an interior point of K, then $x + V \subset K$ for a convex neighborhood V of the origin. Due to the continuity of the map $\lambda \to \lambda x$, there must exist a $\lambda \in (0,1)$ with $\lambda x \in V$. Then $x + \lambda x \in x + V \subset K$, so that $p_K(x) < 1$. Conversely, if $p_K(x) < 1$, then $x \in \lambda K$ for some $\lambda \in (0,1)$, which yields

$$x \in x + (1-\lambda)K \subset \lambda K + (1-\lambda)K = K.$$

Notice that $(1-\lambda)K$ is a neighborhood of the origin.

Since A contains no interior point of K it follows that $f_0(x) = 1 \leq p_K(x)$ for all $x \in A$. If $x \in A$ and $\lambda > 0$, then $f_0(\lambda x) \leq p_K(\lambda x)$, while for $\lambda \leq 0$ we have $f_0(\lambda x) \leq 0 \leq p_K(\lambda x)$. Consequently $f_0 \leq p_K$ on E_0. By Theorem A.1.1, f_0 has a linear extension f to E such that $f \leq p_K$. Put $H = \{x \mid f(x) = 1\}$. Then H is a hyperplane. Since $|f(x)| \leq p_K(x) < 1$ for x in K, it follows that f is bounded on a neighborhood of 0 and thus continuous. Therefore H is a closed hyperplane and it is clear that $H \supset A$ and $H \cap K = \emptyset$. □

Corollary A.2.2 *If U is a nonempty open convex set and F is a linear subspace such that $F \cap U = \emptyset$, then there exists a continuous linear functional f such that $f(x) = 0$ if $x \in F$ and $f(x) > 0$ if $x \in U$.*

In order to prove a strict separation result we need the following lemma of independent interest:

Lemma A.2.3 *Suppose that K_1 and K_2 are two nonempty convex sets in a real linear topological space E with $K_1 \cap K_2 = \emptyset$. If one of them is open, then there exists a closed hyperplane separating K_1 from K_2.*

Proof. If K_1 is open, then the set

$$U = K_1 - K_2 = \bigcup_{k_2 \in K_2} (K_1 - k_2)$$

is open. Since K_1 and K_2 are convex, U is convex too. Moreover, $0 \notin U$ since $K_1 \cap K_2 = \emptyset$. By Corollary A.2.2 there exists a continuous linear functional f such that $f(x) > 0$ on U. Therefore $f(x) > f(y)$ for all $x \in K_1$ and all $y \in K_2$. Letting

$$\alpha = \inf\{f(x) \mid x \in K_1\},$$

one can show immediately that K_1 and K_2 are separated by the closed hyperplane $H = \{x \mid f(x) = \alpha\}$. □

Theorem A.2.4 (Strong separation theorem) *Let K_1 and K_2 be two nonempty convex sets in a real locally convex Hausdorff space E such that $K_1 \cap K_2 = \emptyset$. If K_1 is compact and K_2 is closed, then there exists a closed hyperplane strictly separating K_1 from K_2.*

Particularly, if K is a closed convex set in a locally convex space E and $x \in E$ is not in K, then there exists a functional $f \in E'$ such that

$$f(x) > \sup\{f(y) \mid y \in K\}.$$

Proof. By our hypothesis, there exists an open convex neighborhood W of the origin such that $(K_1 + W) \cap (K_2 + W) = \emptyset$. This follows directly by using *reductio ad absurdum*. Since the sets $K_1 + W$ and $K_2 + W$ are convex and open, from Lemma A.2.3 we infer the existence of a separating hyperplane H. A moment's reflection shows that H separates strictly K_1 from K_2. □

The *closed convex hull* of a subset A of a locally convex space E is the smallest closed convex set $\overline{\text{co}}(A)$ containing A (that is, the intersection of all closed convex sets containing A). From Theorem A.2.4 we can infer the following result on the support of closed convex sets:

Corollary A.2.5 *If A is a nonempty subset of a real locally convex Hausdorff space E, then the closed convex hull $\overline{\text{co}}(A)$ is the intersection of all the closed half-spaces containing A. Equivalently,*

$$\overline{\text{co}}(A) = \bigcap_{f \in E'} \{x \mid f(x) \leq \sup_{y \in A} f(y)\}.$$

This corollary implies:

Corollary A.2.6 *In a real locally convex Hausdorff space E, the closed convex sets and the weakly closed convex sets are the same.*

Finally it is worth mentioning a non-topological version of the separation results above, which is important in optimization theory.

Given a set A in a linear space E, a point a of A is said to be a *core point* if for every $v \in E$, $v \neq a$, there exists an $\varepsilon > 0$ such that $a + \delta v \in A$ for every δ with $|\delta| < \varepsilon$.

Theorem A.2.7 *Let K and M be two nonempty convex sets in a real linear space E. If K contains core points and M contains no core point of K, then K and M can be separated by a hyperplane.*

The details can be easily filled out by adapting the argument given in the topological case.

A.3 The Krein–Milman Theorem

In Section 3.3 we proved that every compact convex set in $\mathbb{R}^n$ is the convex hull of its extreme points. This result can be extended to a very general setting.

Theorem A.3.1 *Let E be a locally convex Hausdorff space and K be a nonempty compact convex subset of E. If U is an open convex subset of K such that* $\operatorname{Ext} K \subset U$, *then $U = K$.*

Proof. Suppose that $U \neq K$ and consider the family $\mathcal{U}$ of all open convex sets in K which are not equal to K. By Zorn's lemma, each set $U \in \mathcal{U}$ is contained in a maximal element V of $\mathcal{U}$.

For each $x \in K$ and $t \in [0,1]$, let $\varphi_{x,t} \colon K \to K$ be the continuous map defined by $\varphi_{x,t}(y) = ty + (1-t)x$.

Assuming $x \in V$ and $t \in [0,1)$, we shall show that $\varphi_{x,t}^{-1}(V)$ is an open convex set which contains V properly, hence $\varphi_{x,t}^{-1}(V) = K$. In fact, this is clear when $t = 0$. If $t \in (0,1)$, then $\varphi_{x,t}$ is a homeomorphism and $\varphi_{x,t}^{-1}(V)$ is an open convex set in K. Moreover,

$$\varphi_{x,t}(\overline{V}) \subset V,$$

which yields $\overline{V} \subset \varphi_{x,t}^{-1}(V)$, hence $\varphi_{x,t}^{-1}(V) = K$ by the maximality of V. Therefore $\varphi_{x,t}(K) \subset V$. For any open convex set W in K the intersection $V \cap W$ is also open and convex, and the maximality of V yields that either $V \cup W = V$ or $V \cup W = K$. In conclusion $K \backslash V$ is precisely a singleton $\{e\}$. But such a point is necessarily an extreme point of K, which is a contradiction. $\square$

Corollary A.3.2 (Krein–Milman theorem) *Let K be a nonempty compact convex subset of a locally convex Hausdorff space E. Then K is the closed convex hull of* $\operatorname{Ext} K$.

Proof. By Theorem A.2.4, the set $L = \overline{\mathrm{co}}(\operatorname{Ext} K)$ is the intersection of all open convex sets containing L. If U is an open subset of K and $U \supset L$, then $U \supset \operatorname{Ext} K$. Hence $U = K$ and $L = K$. $\square$

The above proof of the Krein–Milman theorem yields the existence of extreme points as a consequence of the formula $K = \overline{\mathrm{co}}(\operatorname{Ext} K)$. However this can be checked directly. Call a subset A of K *extremal* if it is closed, nonempty and verifies the following property:

$$x, y \in K \quad \text{and} \quad (1-\lambda)x + \lambda y \in A \quad \text{for some } \lambda \in (0,1) \implies x, y \in A.$$

By Zorn's lemma we can choose a minimal extremal subset, say S. We show that S is a singleton (which yields an extreme point of K). In fact, if S contains more than one point, the separation Theorem A.2.4 proves the existence of a functional $f \in E'$ which is not constant on S. But in this case the set

$$S_0 = \{x \in S \mid f(x) = \sup_{y \in S} f(y)\}$$

will contradict the minimality of S. Now the formula $K = \overline{\mathrm{co}}(\operatorname{Ext} K)$ can easily be proved by noticing that the inclusion $\overline{\mathrm{co}}(\operatorname{Ext} K) \subset K$ cannot be strict.

Another application of Theorem A.3.1 is the following generalization of Theorem 3.4.6:

Corollary A.3.3 (Bauer maximum principle) *Suppose that K is a nonempty compact convex set as in Theorem* A.3.1*. Then every upper semicontinuous convex function $f\colon K \to [-\infty, \infty)$ attains its supremum at an extreme point.*

Proof. Since f is upper semicontinuous, the family of sets

$$U_n = \{x \in K \mid f(x) < n\} \quad (n \in \mathbb{N}),$$

provides an open covering of K, so $K = U_n$ for some n, which shows that f is bounded above. Put $M = \sup\{f(x) \mid x \in K\}$. If f does not attain its supremum at a point of $\operatorname{Ext} K$, then $U = \{x \in K \mid f(x) < M\}$ is an open convex set containing $\operatorname{Ext} K$. By Theorem A.3.1 we conclude that $U = K$, which is a contradiction. $\square$

It is interesting to note the following converse to Theorem A.3.1:

Theorem A.3.4 (D. P. Milman) *Suppose that K is a compact convex set* (*in a locally convex Hausdorff space E*) *and C is a subset of K such that K is the closed convex hull of C. Then the extreme points of K are contained in the closure of C.*

Coming back to Theorem 3.3.5, the fact that every point x of a compact convex set K in $\mathbb{R}^n$ is a convex combination of extreme points of K,

$$x = \sum_{k=1}^{m} \lambda_k x_k,$$

can be reformulated as an integral representation,

$$f(x) = \sum_{k=1}^{m} \lambda_k f(x_k) = \int_{\operatorname{Ext} K} f \, d\mu \tag{A.1}$$

for all $f \in (\mathbb{R}^n)'$. Here $\mu = \sum_{k=1}^{m} \lambda_k \delta_{x_k}$ is a convex combination of Dirac measures δ_{x_k} and thus μ itself is a Borel probability measure on $\operatorname{Ext} K$.

The integral representation (A.1) can be extended to all Borel probability measures μ on a compact convex set K (in a locally convex Hausdorff space E). We shall need some definitions.

Given a Borel probability measure μ on K, and a Borel subset $S \subset K$, we say that μ *is concentrated on* S if $\mu(K \backslash S) = 0$. For example, a Dirac measure δ_x is concentrated on x.

A point $x \in K$ is said to be the *barycenter* of μ provided that

$$f(x) = \int_K f \, d\mu \quad \text{for all } f \in E'.$$

Since the functionals separate the points of E, the point x is uniquely determined by μ. With this preparation, we can reformulate the Krein–Milman theorem as follows:

Theorem A.3.5 *Every point of a compact convex subset K (of a locally convex Hausdorff space E), is the barycenter of a Borel probability measure on K, which is supported by the closure of the extreme points of K.*

H. Bauer pointed out that the extremal points of K are precisely the points $x \in K$ for which the only Borel probability measure μ which admits x as a barycenter is δ_x. See [200, p. 6]. This fact together with Theorem A.3.5 yields D. P. Milman's aforementioned converse of the Krein–Milman theorem. For an alternative argument see [64, pp. 103–104].

Theorem A.3.5 led G. Choquet [53] to his theory on integral representation for elements of a closed convex cone.

B

Elementary Symmetric Functions

The *elementary symmetric functions* of n variables are defined by

$$\begin{aligned}
e_0(x_1, x_2, \ldots, x_n) &= 1 \\
e_1(x_1, x_2, \ldots, x_n) &= x_1 + x_2 + \cdots + x_n \\
e_2(x_1, x_2, \ldots, x_n) &= \sum_{i<j} x_i x_j \\
&\vdots \\
e_n(x_1, x_2, \ldots, x_n) &= x_1 x_2 \cdots x_n.
\end{aligned}$$

The different e_k being of different degrees, they are not comparable. However, they are connected by nonlinear inequalities. To state them, it is more convenient to consider their averages,

$$E_k(x_1, x_2, \ldots, x_n) = e_k(x_1, x_2, \ldots, x_n) \big/ \binom{n}{k}$$

and to write E_k for $E_k(x_1, x_2, \ldots, x_n)$ in order to avoid excessively long formulæ.

B.1 Newton's Inequalities

The simplest set of inequalities relating the elementary symmetric functions was discovered by I. Newton [175] and C. Maclaurin [149]:

Theorem B.1.1 *Let $\mathcal{F}$ be an n-tuple of nonnegative numbers. Then:*

$$E_k^2(\mathcal{F}) > E_{k-1}(\mathcal{F}) \cdot E_{k+1}(\mathcal{F}), \quad 1 \leq k \leq n-1 \tag{N}$$

and

$$E_1(\mathcal{F}) > E_2^{1/2}(\mathcal{F}) > \cdots > E_n^{1/n}(\mathcal{F}) \tag{M}$$

unless all entries of $\mathcal{F}$ coincide.

Actually Newton's inequalities (N) work for n-tuples of real, not necessarily positive, elements. An analytic proof along Maclaurin's ideas will be presented below. In Section B.2 we shall indicate an alternative argument, based on mathematical induction, which yields more Newton type inequalities, in an interpolative scheme.

The inequalities (M) can be deduced from (N) since

$$(E_0E_2)(E_1E_3)^2(E_2E_4)^3\cdots(E_{k-1}E_{k+1})^k < E_1^2E_2^4E_3^6\cdots E_k^{2k}$$

gives $E_{k+1}^k < E_k^{k+1}$ or, equivalently,

$$E_k^{1/k} > E_{k+1}^{1/(k+1)}.$$

Among the inequalities noticed above, the most notable is of course the AM–GM inequality:

$$\left(\frac{x_1+x_2+\cdots+x_n}{n}\right)^n \geq x_1x_2\cdots x_n$$

for all $x_1, x_2, \ldots, x_n \geq 0$. A hundred years after C. Maclaurin, A.-L. Cauchy [50] gave his beautiful inductive argument. Notice that the AM–GM inequality was known to Euclid [73] in the special case where $n = 2$.

Remark B.1.2 Newton's inequalities were intended to solve the problem of counting the number of imaginary roots of an algebraic equation. In Chapter 2 of Part 2 of *Arithmetica Universalis*, entitled *De Forma Æquationis*, Newton made (without proof) the following statement: *Given an equation with real coefficients,*

$$a_0x^n + a_1x^{n-1} + \cdots + a_n = 0 \quad (a_0 \neq 0),$$

the number of its imaginary roots cannot be less than the number of changes of sign that occur in the sequence

$$a_0^2,\ \left(\frac{a_1}{\binom{n}{1}}\right)^2 - \frac{a_2}{\binom{n}{2}}\cdot\frac{a_0}{\binom{n}{0}},\ \ldots,\ \left(\frac{a_{n-1}}{\binom{n}{n-1}}\right)^2 - \frac{a_n}{\binom{n}{n}}\cdot\frac{a_{n-2}}{\binom{n}{n-2}},\ a_n^2.$$

Accordingly, if all the roots are real, then all the entries in the above sequence must be nonnegative (a fact which yields Newton's inequalities).

Trying to understand which was Newton's argument, C. Maclaurin [149] gave a direct proof of the inequalities (N) and (M), but the Newton counting problem remained open until 1865, when J. Sylvester [234, 235] succeeded in proving a remarkable general result.

Quite unexpectedly, it is the real algebraic geometry (not analysis) which gives us the best understanding of Newton's inequalities. The basic fact (discovered by J. Sylvester) concerns the *semi-algebraic character* of the set of all real polynomials with all roots real:

Theorem B.1.3 (J. Sylvester) *For each natural number $n \geq 2$ there exists a set of at most $n-1$ polynomials with integer coefficients,*

$$R_{n,1}(x_1, \ldots, x_n), \ldots, R_{n,k(n)}(x_1, \ldots, x_n), \tag{R_n}$$

such that the monic real polynomials of order n,

$$P(x) = x^n + a_1 x^{n-1} + \cdots + a_n,$$

which have only real roots are precisely those for which

$$R_{n,1}(a_1, \ldots, a_n) \geq 0, \ldots, R_{n,k(n)}(a_1, \ldots, a_n) \geq 0.$$

The above result can be seen as a generalization of the well-known fact that *the roots of a quadratic polynomial $x^2 + a_1 x + a_2$ are real if and only if its discriminant*

$$D_2(1, a_1, a_2) = a_1^2 - 4a_2 \tag{D_2}$$

is nonnegative.

Theorem B.1.3 is built on the Sturm method of counting real roots, taking into account that only the leading coefficients enter into play. It turns out that they are nothing but the principal subresultant coefficients (with convenient signs added), which are determinants extracted from the Sylvester matrix.

A set $(R_{n,k})_k^{k(n)}$ as in Theorem B.1.3 will be called a *Sylvester family* (of order n).

In Sylvester's approach, $R_{n,1}(a_1, \ldots, a_n)$ equals the *discriminant* D_n of the polynomial $P(x) = x^n + a_1 x^{n-1} + \cdots + a_n$, that is,

$$D_n = D_n(1, a_1, \ldots, a_n) = \prod_{1 \leq i < j \leq n} (x_i - x_j)^2,$$

where $x_1, \ldots, x_n$ are the roots of $P(x)$; D_n is a symmetric and homogeneous (of degree $n^2 - n$) polynomial in $\mathbb{Z}[x_1, \ldots, x_n]$. For details, see [19]. Unfortunately, at present no compact formula for D_n is known. According to [220], the number of nonzero coefficients in the expression for the discriminant increases rapidly with the degree; e.g., D_9 has 26095 terms.

For $n \in \{2, 3\}$ one can indicate Sylvester families consisting of just a single polynomial, the corresponding discriminant. An inspection of the argument given by L. Euler to solve in radicals the quartic equations allows us to write down a Sylvester family for $n = 4$. See the paper by C. P. Niculescu [177].

Remark B.1.4 Given a Sylvester family for $n = N$, we can easily indicate such a family for each $n \in \{1, \ldots, N\}$; the trick is to replace a $P(x)$ of degree n by $x^{N-n}P(x)$, which is of degree N.

Also, any Sylvester family $(R_{n,k})_{k=1}^{k(n)}$ (for some $n \geq 2$), allows us to decide which monic real polynomial $P(x) = x^n + a_1 x^{n-1} + \cdots + a_n$ has only nonnegative roots. A set of (necessary and) sufficient conditions consists of the following inequalities:

$$-a_1 \geq 0, \ldots, (-1)^n a_n \geq 0$$

and

$$R_{n,1}(a_1, \ldots, a_n) \geq 0, \ldots, R_{n,k(n)}(a_1, \ldots, a_n) \geq 0.$$

In fact, under the above circumstances, $x < 0$ yields $P(x) \neq 0$.

The Newton inequalities (N) were proved in [99] following Maclaurin's argument. The basic ingredient is the following lemma, a consequence of repeated application of Rolle's theorem, which we give here under the formulation of J. Sylvester [235]:

Lemma B.1.5 *If*

$$F(x, y) = c_0 x^n + c_1 x^{n-1} y + \cdots + c_n y^n$$

is a homogeneous function of the n-th degree in x and y, which has all its roots x/y real, then the same is true for all non-identical 0 *equations*

$$\frac{\partial^{i+j} F}{\partial x^i \partial y^j} = 0,$$

obtained from it by partial differentiation with respect to x and y. Further, if E is one of these equations, and it has a multiple root α, then α is also a root, of multiplicity one higher, of the equation from which E is derived by differentiation.

Any polynomial of the n-th degree, with real roots, can be represented as

$$E_0 x^n - \binom{n}{1} E_1 x^{n-1} + \binom{n}{2} E_1 x^{n-2} - \cdots + (-1)^n E_n$$

and we shall apply Lemma B.1.5 to the associated homogeneous polynomial

$$F(x, y) = E_0 x^n - \binom{n}{1} E_1 x^{n-1} y + \binom{n}{2} E_1 x^{n-2} y^2 - \cdots + (-1)^n E_n y^n.$$

Considering the case of the derivatives

$$\frac{\partial^{n-2} F}{\partial x^k \partial y^{n-2-k}} \quad (\text{for } k = 0, \ldots, n-2),$$

we arrive at the fact that all the quadratic polynomials

$$E_{k-1} x^2 - 2E_k xy + E_{k+1} y^2$$

for $k = 0, \ldots, n-2$ also have real roots. Consequently, the Newton inequalities express precisely this fact in the language of discriminants. That is why we shall refer to (N) as the *quadratic Newton inequalities.*

Stopping a step ahead, we get what S. Rosset [217] called the *cubic Newton inequalities*:

$$6E_kE_{k+1}E_{k+2}E_{k+3} + 3E_{k+1}^2E_{k+2}^2 \geq 4E_kE_{k+2}^3 + E_k^2E_{k+3}^2 + 4E_{k+1}^3E_{k+3} \quad (\mathrm{N}_3)$$

for $k = 0, \ldots, n-3$. They are motivated by the well-known fact that a cubic real polynomial

$$x^3 + a_1x^2 + a_2x + a_3$$

has only real roots if and only if its discriminant

$$\begin{aligned} D_3 &= D_3(1, a_1, a_2, a_3) \\ &= 18a_1a_2a_3 + a_1^2a_2^2 - 27a_3^2 - 4a_2^3 - 4a_1^3a_3 \end{aligned}$$

is nonnegative. Consequently, the equation

$$E_kx^3 - 3E_{k+1}x^2y + 3E_{k+2}xy^2 - E_{k+3}y^3 = 0$$

has all its roots x/y real if and only if (N_3) holds.

S. Rosset [217] derived the inequalities (N_3) by an inductive argument and noticed that they are strictly stronger than (N). In fact, (N_3) can be rewritten as

$$4(E_{k+1}E_{k+3} - E_{k+2}^2)(E_kE_{k+2} - E_{k+1}^2) \geq (E_{k+1}E_{k+2} - E_kE_{k+3})^2$$

which yields (N).

As concerns the Newton inequalities (N_n) of order $n \geq 2$ (when applied to strings of $m \geq n$ elements), they consist of at most $n-1$ sets of relations, the first one being

$$D_n\left(1, (-1)^1\binom{n}{1}\frac{E_{k+1}}{E_k}, (-1)^2\binom{n}{2}\frac{E_{k+2}}{E_k}, \ldots, (-1)^n\binom{n}{n}\frac{E_{k+n}}{E_k}\right) \geq 0$$

for $k \in \{0, \ldots, m-n\}$.

Notice that each of these inequalities is homogeneous (for example, the last one consists of terms of weight $n^2 - n$) and the sum of all coefficients in the left hand side is 0.

B.2 More Newton Inequalities

Our argument will yield a bit more, precisely the log concavity of the functions $E_k \colon k \to E_k(\mathcal{F})$:

Theorem B.2.1 *Suppose that $\alpha, \beta \in \mathbb{R}_+$ and $j, k \in \mathbb{N}$ are numbers such that*

$$\alpha + \beta = 1 \quad \textit{and} \quad j\alpha + k\beta \in \{0, \ldots, n\}.$$

Then

$$E_{j\alpha+k\beta}(\mathcal{F}) \geq E_j^\alpha(\mathcal{F}) \cdot E_k^\beta(\mathcal{F}),$$

for every n-tuple $\mathcal{F}$ of nonnegative real numbers. Moreover, equality occurs if and only if all entries of $\mathcal{F}$ are equal.

The proof will be done by induction on the length of $\mathcal{F}$.

According to Rolle's theorem, if all roots of a polynomial $P \in \mathbb{R}[X]$ are real (respectively, real and distinct), then the same is true for its derivative P'. Given an n-tuple $\mathcal{F} = (x_1, \ldots, x_n)$, we shall attach to it the polynomial

$$P_{\mathcal{F}}(x) = (x - x_1) \cdots (x - x_n) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} E_k(x_1, \ldots, x_n) x^{n-k}.$$

The $(n-1)$-tuple $\mathcal{F}' = \{y_1, \ldots, y_{n-1}\}$, consisting of all roots of the derivative of $P_{\mathcal{F}}(x)$ will be called the *derived n-tuple* of $\mathcal{F}$. Because

$$(x - y_1) \cdots (x - y_{n-1}) = \sum_{k=0}^{n-1} (-1)^k \binom{n-1}{k} E_k(y_1, \ldots, y_{n-1}) x^{n-k}$$

and

$$\begin{aligned} (x - y_1) \cdots (x - y_{n-1}) &= \frac{1}{n} \cdot \frac{dP_{\mathcal{F}}}{dx}(x) \\ &= \sum_{k=0}^{n} (-1)^k \, \frac{n-k}{n} \binom{n}{k} E_k(x_1, \ldots, x_n) x^{n-k-1} \\ &= \sum_{k=0}^{n-1} (-1)^k \binom{n-1}{k} E_k(x_1, \ldots, x_n) x^{n-1-k} \end{aligned}$$

we are led to the following result, which enables us to reduce the number of variables when dealing with symmetric functions.

Lemma B.2.2 $E_j(\mathcal{F}) = E_j(\mathcal{F}')$ *for every* $j \in \{0, \ldots, |\mathcal{F}| - 1\}$.

Another simple but useful fact is the following:

Lemma B.2.3 *Suppose that $\mathcal{F}$ is an n-tuple of real numbers and $0 \notin \mathcal{F}$. Put $\mathcal{F}^{-1} = \{1/a \mid a \in \mathcal{F}\}$. Then*

$$E_j(\mathcal{F}^{-1}) = E_{n-j}(\mathcal{F}) / E_n(\mathcal{F})$$

for every $j \in \{0, \ldots, n\}$.

Proof of Theorem B.2.1. For $|\mathcal{F}| = 2$ we have to prove just one inequality, namely, $x_1 x_2 \leq (x_1 + x_2)^2/4$, which is clearly valid for every $x_1, x_2 \in \mathbb{R}$; the equality occurs if and only if $x_1 = x_2$.

Suppose now that the assertion of Theorem B.2.1 holds for all k-tuples with $k \leq n - 1$. Let $\mathcal{F}$ be a n-tuple of nonnegative numbers ($n \geq 3$), let $j, k \in \mathbb{N}$, and $\alpha, \beta \in \mathbb{R}_+ \backslash \{0\}$ be numbers such that

$$\alpha + \beta = 1 \quad \text{and} \quad j\alpha + k\beta \in \{0, \ldots, n\}.$$

According to Lemma B.2.2 (and our inductive hypothesis), we have

$$E_{j\alpha+k\beta}(\mathcal{F}) \geq E_j^\alpha(\mathcal{F}) \cdot E_k^\beta(\mathcal{F}),$$

except for the case where $j < k = n$ or $k < j = n$. Suppose, for example, that $j < k = n$; then necessarily $j\alpha + n\beta < n$. We have to show that

$$E_{j\alpha+n\beta}(\mathcal{F}) \geq E_j^\alpha(\mathcal{F}) \cdot E_n^\beta(\mathcal{F}).$$

If $0 \in \mathcal{F}$, then $E_n(\mathcal{F}) = 0$, and the inequality is clear; the equality occurs if and only if $E_{j\alpha+n\beta}(\mathcal{F}') = E_{j\alpha+n\beta}(\mathcal{F}) = 0$, that is (according to our inductive hypothesis), when all entries of $\mathcal{F}$ coincide.

If $0 \notin \mathcal{F}$, then by Lemma B.2.3 we have to prove that

$$E_{n-j\alpha-n\beta}(\mathcal{F}^{-1}) \geq E_{n-j}^\alpha(\mathcal{F}^{-1}),$$

or, equivalently (see Lemma B.2.2), $E_{n-j\alpha-n\beta}((\mathcal{F}^{-1})') \geq E_{n-j}^\alpha((\mathcal{F}^{-1})')$, which is true by our hypothesis. $\square$

Notice that the argument above covers Newton's inequalities even for n-tuples of real (not necessarily positive) elements.

The general problem of comparing monomials in $E_1, \ldots, E_n$ was completely solved by G. H. Hardy, J. E. Littlewood and G. Pólya in [99, Theorem 77, p. 64]:

Theorem B.2.4 *Let $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n$ be nonnegative numbers. Then*

$$E_1^{\alpha_1}(\mathcal{F}) \cdots E_n^{\alpha_n}(\mathcal{F}) \leq E_1^{\beta_1}(\mathcal{F}) \cdots E_n^{\beta_n}(\mathcal{F})$$

for every n-tuple $\mathcal{F}$ of positive numbers if and only if

$$\alpha_m + 2\alpha_{m+1} + \cdots + (n-m+1)\alpha_n \geq \beta_m + 2\beta_{m+1} + \cdots + (n-m+1)\beta_n$$

for $1 \leq m \leq n$, with equality when $m = 1$.

An alternative proof, also based on Newton's inequalities (N), is given in [155, p. 93].

B.3 A Result of H. F. Bohnenblust

The elementary symmetric functions e_r are positively homogeneous of degree r and verify a property of concavity which was already noticed in the case of $e_n = x_1 \cdots x_n$ (see Section 3.9, Exercise 2):

Theorem B.3.1 (Bohnenblust's inequality) *The sum of two n-tuples of nonnegative numbers $\mathcal{F} = \{x_1, \ldots, x_n\}$ and $\mathcal{G} = \{y_1, \ldots, y_n\}$ is defined by the formula $\mathcal{F} + \mathcal{G} = \{x_1 + y_1, \ldots, x_n + y_n\}$. Then*

$$e_r(\mathcal{F} + \mathcal{G})^{1/r} \geq e_r(\mathcal{F})^{1/r} + e_r(\mathcal{G})^{1/r}$$

for every $r = 1, \ldots, n$. Moreover, the equality occurs only when the entries of $\mathcal{F}$ and $\mathcal{G}$ are proportional.

In other words, the functions $e_r(\mathcal{F})^{1/r}$ are strictly concave (as functions of $x_1, \ldots, x_n$).

The argument given here is due to M. Marcus and J. Lopes [154]. It combines a special case of Minkowski's inequality with the following lemma:

Lemma B.3.2 (Marcus–Lopes inequality) *Under the hypotheses of Theorem* B.3.1*, for $r = 1, \ldots, n$ and n-tuples of nonnegative numbers not all zero, we have*

$$\frac{e_r(\mathcal{F}+\mathcal{G})}{e_{r-1}(\mathcal{F}+\mathcal{G})} \geq \frac{e_r(\mathcal{F})}{e_{r-1}(\mathcal{F})} + \frac{e_r(\mathcal{G})}{e_{r-1}(\mathcal{G})}.$$

The inequality is strict unless $r = 1$ or there exists a $\lambda > 0$ such that $\mathcal{F} = \lambda\mathcal{G}$.

Proof. For $r = 1$ the inequality is actually an equality. For $r = 2$, we have to look at the following identity:

$$\frac{e_2(\mathcal{F}+\mathcal{G})}{e_1(\mathcal{F}+\mathcal{G})} - \frac{e_2(\mathcal{F})}{e_1(\mathcal{F})} - \frac{e_2(\mathcal{G})}{e_1(\mathcal{G})} = \frac{\sum_{k=1}^n (x_k \sum_{j=1}^n y_j - y_k \sum_{j=1}^n x_j)^2}{2e_1(\mathcal{F}+\mathcal{G})\, e_1(\mathcal{F})\, e_1(\mathcal{G})}.$$

Assume now that $r > 2$. For an *n-tuple* $\mathcal{H} = \{z_1, \ldots, z_n\}$ we shall denote $\mathcal{H}_{\hat{k}} = \{z_1, \ldots, \widehat{z_k}, \ldots, z_n\}$, where the cap indicates omission. Then:

$$\sum_{k=1}^n x_k e_{r-1}(\mathcal{F}_{\hat{k}}) = r e_r(\mathcal{F}) \tag{B.1}$$

$$x_k e_{r-1}(\mathcal{F}_{\hat{k}}) + e_r(\mathcal{F}_{\hat{k}}) = e_r(\mathcal{F}). \tag{B.2}$$

Summing on k in (B.2) we obtain

$$n e_r(\mathcal{F}) = \sum_{k=1}^n x_k e_{r-1}(\mathcal{F}_{\hat{k}}) + \sum_{k=1}^n e_r(\mathcal{F}_{\hat{k}}),$$

and thus from (B.1) we infer that $\sum_{k=1}^n e_r(\mathcal{F}_{\hat{k}}) = (n-r)e_r(\mathcal{F})$. Since

$$\begin{aligned} e_r(\mathcal{F}) - e_r(\mathcal{F}_{\hat{k}}) &= x_k e_{r-1}(\mathcal{F}_{\hat{k}}) \\ &= x_k e_{r-1}(\mathcal{F}) - x_k^2 e_{r-2}(\mathcal{F}_{\hat{k}}) \end{aligned}$$

we obtain

$$r e_r(\mathcal{F}) = \sum_{k=1}^n x_k e_{r-1}(\mathcal{F}) - \sum_{k=1}^n x_k^2 e_{r-2}(\mathcal{F}_{\hat{k}})$$

and thus

$$\begin{aligned} \frac{e_r(\mathcal{F})}{e_{r-1}(\mathcal{F})} &= \frac{1}{r}\left[\sum_{k=1}^n x_k - \sum_{k=1}^n \frac{x_k^2\, e_{r-2}(\mathcal{F}_{\hat{k}})}{e_{r-1}(\mathcal{F})}\right] \\ &= \frac{1}{r}\left[\sum_{k=1}^n x_k - \sum_{k=1}^n \frac{x_k^2}{x_k + e_{r-1}(\mathcal{F}_{\hat{k}})/e_{r-2}(\mathcal{F}_{\hat{k}})}\right]. \end{aligned}$$

Therefore

$$\begin{aligned}\Delta &= \frac{e_r(\mathcal{F}+\mathcal{G})}{e_{r-1}(\mathcal{F}+\mathcal{G})} - \frac{e_r(\mathcal{F})}{e_{r-1}(\mathcal{F})} - \frac{e_r(\mathcal{G})}{e_{r-1}(\mathcal{G})}\\ &= \frac{1}{r}\sum_{k=1}^{n}\Big[\frac{x_k^2}{x_k + f_{r-1}(\mathcal{F}_{\hat{k}})} + \frac{y_k^2}{y_k + f_{r-1}(\mathcal{G}_{\hat{k}})} - \frac{(x_k+y_k)^2}{x_k+y_k+f_{r-1}((\mathcal{F}+\mathcal{G})_{\hat{k}})}\Big]\end{aligned}$$

where $f_s(\mathcal{F}) = e_s(\mathcal{F})/e_{s-1}(\mathcal{F})$.

The proof ends by induction. Assume that the statement of the theorem is true for $r-1$, that is,

$$f_{r-1}((\mathcal{F}+\mathcal{G})_{\hat{k}}) > f_{r-1}(\mathcal{F}_{\hat{k}}) + f_{r-1}(\mathcal{G}_{\hat{k}}), \tag{B.3}$$

unless $\mathcal{F}_{\hat{k}}$ and $\mathcal{G}_{\hat{k}}$ are proportional (when equality holds). Then

$$\begin{aligned}\Delta &> \frac{1}{r}\sum_{k=1}^{n}\Big[\frac{x_k^2}{x_k + f_{r-1}(\mathcal{F}_{\hat{k}})} + \frac{y_k^2}{y_k + f_{r-1}(\mathcal{G}_{\hat{k}})}\\ &\qquad\qquad - \frac{(x_k+y_k)^2}{x_k+y_k+f_{r-1}(\mathcal{F}_{\hat{k}}) + f_{r-1}(\mathcal{G}_{\hat{k}})}\Big]\\ &= \frac{1}{r}\sum_{k=1}^{n}\frac{[x_k f_{r-1}(\mathcal{G}_{\hat{k}}) - y_k f_{r-1}(\mathcal{F}_{\hat{k}})]^2}{[x_k + f_{r-1}(\mathcal{F}_{\hat{k}})][y_k + f_{r-1}(\mathcal{G}_{\hat{k}})][x_k + y_k + f_{r-1}(\mathcal{F}_{\hat{k}}) + f_{r-1}(\mathcal{G}_{\hat{k}})]}\end{aligned}$$

provided that at least one of the inequalities (B.3) is strict. □

Proof of Theorem B.3.1. In fact, by Minkowski's inequality for $p = 0$ and Lemma B.3.2, we have

$$\begin{aligned}e_r(\mathcal{F}+\mathcal{G})^{1/r} &= \Big[\frac{e_r(\mathcal{F}+\mathcal{G})}{e_{r-1}(\mathcal{F}+\mathcal{G})}\cdot\frac{e_{r-1}(\mathcal{F}+\mathcal{G})}{e_{r-2}(\mathcal{F}+\mathcal{G})}\cdots\frac{e_1(\mathcal{F}+\mathcal{G})}{e_0(\mathcal{F}+\mathcal{G})}\Big]^{1/r}\\ &\geq \Big\{\Big[\frac{e_r(\mathcal{F})}{e_{r-1}(\mathcal{F})} + \frac{e_r(\mathcal{G})}{e_{r-1}(\mathcal{G})}\Big]\cdots\Big[\frac{e_1(\mathcal{F})}{e_0(\mathcal{F})} + \frac{e_1(\mathcal{G})}{e_0(\mathcal{G})}\Big]\Big\}^{1/r}\\ &\geq \Big(\prod_{k=1}^{r}\frac{e_k(\mathcal{F})}{e_{k-1}(\mathcal{F})}\Big)^{1/r} + \Big(\prod_{k=1}^{r}\frac{e_k(\mathcal{G})}{e_{k-1}(\mathcal{G})}\Big)^{1/r}\\ &= e_r(\mathcal{F})^{1/r} + e_r(\mathcal{G})^{1/r}.\end{aligned}$$

The problem of equality is left to the reader. □

Bohnenblust's inequality has important consequences to positive matrices $A \in \mathrm{M}_n(\mathbb{C})$, $A = (a_{ij})_{i,j=1}^{n}$. In this case all eigenvalues $\lambda_1(A), \ldots, \lambda_n(A)$ are nonnegative and the symmetric elementary functions of them can be easily computed via the Cauchy–Binet formulae:

$$\sum_{k=1}^{n} \lambda_k(A) = \sum_{k=1}^{n} a_{kk}$$

$$\sum_{i<j} \lambda_i(A)\lambda_j(A) = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \cdots + \det \begin{pmatrix} a_{n-1\,n-1} & a_{n-1\,n} \\ a_{n\,n-1} & a_{n\,n} \end{pmatrix}$$

$$\vdots$$

$$\prod_{k=1}^{n} \lambda_k(A) = \det(a_{ij})_{i,j=1}^{n}.$$

As a consequence, Theorem B.3.1 implies the result of Exercise 4 in Section 3.4: If A, B are positive matrices, and $\alpha \in (0,1)$, then

$$\begin{aligned} \big(\det((1-\alpha)A + \alpha B)\big)^{1/n} &\geq (1-\alpha)(\det A)^{1/n} + \alpha(\det B)^{1/n} \\ &\geq (\det A)^{(1-\alpha)/n}(\det B)^{\alpha/n}. \end{aligned}$$

Newton's inequalities (as well as all Newton inequalities of higher order) have equivalent formulations in terms of positive matrices (and their principal minors). We shall recall here the analogue of the AM–GM inequality: *If A is a positive matrix in* $\mathrm{M}_n(\mathbb{R})$, *then*

$$\left(\frac{\operatorname{Trace} A}{n}\right)^n > \det A,$$

unless A is a multiple of the identity I.

In differential geometry, the *higher-order mean curvatures* are defined as the elementary symmetric functions of the principal curvatures. In fact, if S is a hypersurface in $\mathbb{R}^n$ and p is a point of S, one considers the Gauss map, $g\colon p \to N(p)$, whose differential at p is diagonalized by the principal curvature directions at p,

$$dg_p(e_j) = -k_j e_j \quad \text{for } j = 1, \ldots, n.$$

Then the j-th-order mean curvatures H_j are given by

$$\prod_{k=1}^{n-1}(1 + tk_j) = \sum_{j=0}^{n-1} \binom{n-1}{j} H_j t^j.$$

See R. Osserman [193] for details. It would be interesting to explore the applications of various inequalities of convexity to this area.

C

The Variational Approach of PDE

The aim of this appendix is to illustrate a number of problems in partial differential equations (PDE) which can be solved by seeking a global minimum of suitable convex functionals. This idea goes back to advanced calculus. See the comments at the end of Section 3.8.

C.1 The Minimum of Convex Functionals

The main criterion for the existence and uniqueness of global minimum of convex functions is actually a far reaching generalization of the orthogonal projection:

Theorem C.1.1 *Let C be a closed convex set in a reflexive Banach space V and let $J\colon C \to \mathbb{R}$ be a convex function such that:*

(i) *J is weakly lower semicontinuous, that is,*

$$u_n \to u \text{ weakly in } V \text{ implies } J(u) \leq \liminf_{n\to\infty} J(u_n);$$

(ii) *Either C is bounded, or $\lim_{\|u\|\to\infty} J(u) = \infty$.*

Then J admits at least one global minimum and the points of global minimum constitutes a convex set.

If, moreover, J is strictly convex, then there is a unique global minimum.

Proof. Put

$$m = \inf_{u\in C} J(u).$$

Clearly, $m < \infty$, and there exists a sequence $(u_n)_n$ of elements in C such that $J(u_n) \to m$. By our hypotheses, the sequence $(u_n)_n$ is bounded, so by Theorem A.1.6, we may assume (replacing $(u_n)_n$ by a subsequence) that it is also weakly converging to an element u in C. Here we used the fact that C is weakly closed (which is a consequence of Corollary A.2.6). Then

$$m \leq J(u) \leq \liminf_{n\to\infty} J(u_n) = m,$$

and thus u is a global minimum. The remainder of the proof is left to the reader as an exercise. □

In the differentiable case we state the following useful version of Theorem C.1.1:

Theorem C.1.2 *Let V be a reflexive Banach space V and let $J: V \to \mathbb{R}$ be a Gâteaux differentiable convex functional with the following properties:*

(i) *For each $u \in V$, the map $J'(u): v \to J'(u;v)$ is an element of V';*
(ii) $\lim_{\|u\|\to\infty} J(u) = \infty$.

Then J admits at least one global minimum and the points of global minimum are precisely the points u such that

$$J'(u\,;v) = 0 \quad \textit{for all } v \in V.$$

If, moreover, J is strictly convex, then there is a unique global minimum.

Proof. First notice that J is weakly lower semicontinuous. In fact, by Theorem 3.9.1,

$$J(u_n) \geq J(u) + J'(u\,;u_n - u)$$

for all n, while $J'(u\,;u_n - u) = J'(u)(u_n - u) \to 0$ by our hypotheses. Hence, according to Theorem C.1.1, J admits global minima.

If u is a global minimum, then for each $v \in V$ there is a $\delta > 0$ such that

$$\frac{J(u+\varepsilon v) - J(u)}{\varepsilon} \geq 0 \quad \text{whenever } |\varepsilon| < \delta.$$

This yields $J'(u\,;v) \geq 0$. Replacing v by $-v$, we obtain

$$-J'(u\,;v) = J'(u\,;-v) \geq 0,$$

and thus $J'(u\,;v) = 0$. Conversely, if $J'(u\,;v) = 0$ for all $v \in V$, then by Theorem 3.9.1 we get

$$J(v) \geq J(u) + J'(u, v-u) = J(u),$$

that is, u is a global minimum. □

Typically, Theorem C.1.1 applies to functionals of the form

$$J(u) = \frac{1}{2}\,\|u - w\|^2 + \varphi(u), \quad u \in V,$$

where V is an L^p-space with $p \in (1,\infty)$, w is an arbitrary fixed element of V and $\varphi: V \to \mathbb{R}$ is a weakly lower semicontinuous convex function. Theorem C.1.2 covers a large range of well-behaved convex functionals, with important consequences to the problem of existence of solutions of partial differential equations:

Corollary C.1.3 *Let Ω be a nonempty open set in $\mathbb{R}^n$ and let $p > 1$. Consider a function $g \in C^1(\mathbb{R})$ which verifies the following properties:*

(i) $g(0) = 0$ *and* $g(t) \geq \alpha|t|^p$ *for a suitable constant* $\alpha > 0$*;*

(ii) *The derivative* g' *is increasing and* $|g'(t)| \leq \beta|t|^{p-1}$ *for a suitable constant* $\beta > 0$.

Then the linear space $V = L^p(\Omega) \cap L^2(\Omega)$ is reflexive when endowed with the norm

$$\|u\|_V = \|u\|_{L^p} + \|u\|_{L^2},$$

and for all $f \in L^2(\Omega)$ the functional

$$J(u) = \int_\Omega g(u(x))\,dx + \frac{1}{2}\int_\Omega |u(x)|^2\,dx + \int_\Omega f(x)u(x)\,dx, \quad u \in V$$

is convex and Gâteaux differentiable with

$$J'(u;v) = \int_\Omega g'(u(x))v(x)\,dx + \int_\Omega u(x)v(x)\,dx + \int_\Omega f(x)v(x)\,dx.$$

Moreover, J admits a unique global minimum $\bar{u}$, which is the solution of the equation

$$J'(u\,;v) = 0 \quad \text{for all } v \in V.$$

Proof. V is a closed subspace of $L^2(\Omega)$ and thus it is a reflexive space. Then notice that

$$\begin{aligned} |g(t)| &= |g(t) - g(0)| \\ &= \left|\int_0^t g'(s)\,ds\right| \leq \frac{\beta}{p}\,|t|^p, \end{aligned}$$

from which it follows easily that J is well defined. Letting

$$J_1(u) = \int_\Omega g(u(x))\,dx,$$

by Lagrange's mean value theorem,

$$\begin{aligned} J_1(u + tv) &= \int_\Omega g(u(x) + tv(x))\,dx \\ &= \int_\Omega g(u(x))\,dx + t\int_\Omega g'(u(x) + \tau(x)v(x))v(x)\,dx, \end{aligned}$$

where $0 < \tau(x) < t$ for all x, provided that $t > 0$. Then

$$\frac{J_1(u + tv) - J_1(u)}{t} = \int_\Omega g'(u(x) + \tau(x)v(x))v(x)\,dx,$$

and letting $t \to 0+$ we get the desired formula for $J'(u;v)$.

Again by Lagrange's mean value theorem, and the fact that g' is increasing, we have

$$\begin{aligned} J_1(v) &= J_1(u) + \int_\Omega g'(u(x) + \tau(x)(v(x) - u(x))) \cdot (v(x) - u(x))\, dx \\ &\geq J_1(u) + \int_\Omega g'(u(x)) \cdot (v(x) - u(x))\, dx \\ &= J_1(u) + J_1'(u, v - u), \end{aligned}$$

which shows that J_1 is convex. Then the functional J is the sum of a convex function and a strictly convex function.

Finally,

$$\begin{aligned} J(u) &\geq \alpha \int_\Omega |u(x)|^p\, dx + \frac{1}{2} \int_\Omega |u(x)|^2\, dx - \left| \int_\Omega f(x)u(x)\, dx \right| \\ &\geq \alpha \|u\|_{L^p}^p + \frac{1}{2} \|u\|_{L^2}^2 - \|f\|_{L^2} \|u\|_{L^2}, \end{aligned}$$

from which it follows that

$$\lim_{\|u\|_V \to \infty} J(u) = \infty,$$

and the conclusion follows from Theorem C.1.2. □

The result of Corollary C.1.3 extends (with obvious changes) to the case where V is defined as the space of all $u \in L^2(\Omega)$ such that $Au \in L^p(\Omega)$ for a given linear differential operator A. Also, we can consider finitely many functions g_k (verifying the conditions (i) and (ii) for different exponents $p_k > 1$) and finitely many linear differential operators A_k. In that case we shall deal with the functional

$$J(u) = \sum_{k=1}^m \int_\Omega g_k(A_k u)\, dx + \frac{1}{2} \int_\Omega |u|^2\, dx + \int_\Omega f u\, dx,$$

defined on $V = \bigcap_{k=1}^m L^{p_k}(\Omega) \cap L^2(\Omega)$; V is reflexive when endowed with the norm

$$\|u\|_V = \sum_{k=1}^m \|A_k u\|_{L^{p_k}} + \|u\|_{L^2}.$$

C.2 Preliminaries on Sobolev Spaces

Some basic results on Sobolev spaces are recalled here for the convenience of the reader. The details are available from many sources, including [6], [74], [211] and [252].

Let Ω be a bounded open set in $\mathbb{R}^n$ with Lipschitz boundary $\partial\Omega$, and let m be a positive integer.

The Sobolev space $H^m(\Omega)$ consists of all functions $u \in L^2(\Omega)$ which admit *weak derivatives* $D^\alpha u$ in $L^2(\Omega)$, for all multi-indices α with $|\alpha| \leq m$. This means the existence of functions $v_\alpha \in L^2(\Omega)$ such that

$$\int_\Omega v_\alpha \cdot \varphi \, dx = (-1)^{|\alpha|} \int_\Omega u \cdot D^\alpha \varphi \, dx \tag{C.1}$$

for all φ in the space $C_c^\infty(\Omega)$ and all α with $|\alpha| \leq m$. Due to the denseness of $C_c^\infty(\Omega)$ in $L^2(\Omega)$, the functions v_α are uniquely defined by (C.1), and they are usually denoted as $D^\alpha u$.

One can prove easily that $H^m(\Omega)$ is a Hilbert space when endowed with the norm $\|\cdot\|_{H^m}$ associated to the inner product

$$\langle u, v\rangle_{H^m} = \sum_{|\alpha|\leq m} \int_\Omega D^\alpha u \cdot D^\alpha v \, dx.$$

Notice that $C^m(\overline{\Omega})$ is a dense subspace of $H^m(\Omega)$.

Theorem C.2.1 (The trace theorem) *There is a continuous linear operator*

$$\gamma = (\gamma_0, \ldots, \gamma_{m-1}) \colon H^m(\Omega) \to L^2(\partial\Omega)^{m-1}$$

such that

$$\gamma_0 u = u|_{\partial\Omega}, \quad \gamma_1 u = \frac{\partial u}{\partial n}, \ldots, \gamma_{m-1} u = \frac{\partial^{m-1} u}{\partial n^{m-1}}$$

for all u in $C^m(\overline{\Omega})$.

The closure of $C_c^\infty(\Omega)$ in $H^m(\Omega)$ is the Sobolev space $H_0^m(\Omega)$. This space coincides with the kernel of the trace operator γ, indicated in Theorem C.2.1.

On $H_0^1(\Omega)$, the norm $\|\cdot\|_{H^1}$ can be replaced by an equivalent norm,

$$\|u\|_{H_0^1} = \left(\int_\Omega \|\nabla u\|^2 \, dx \right)^{1/2}.$$

In fact, there exists a constant $c > 0$ such that

$$\|u\|_{H_0^1} \leq \|u\|_{H^1} \leq c\|u\|_{H_0^1} \quad \text{for all } u \in H_0^1(\Omega).$$

This is a consequence of a basic inequality in partial differential equations:

Theorem C.2.2 (Poincaré's inequality) *If Ω is a bounded open subset of $\mathbb{R}^n$, then there exists a constant $C > 0$ such that*

$$\|u\|_{L^2} \leq C \left(\int_\Omega \|\nabla u\|^2 \, dx \right)^{1/2}$$

for all $u \in H_0^1(\Omega)$.

Proof. Since $C_c^\infty(\Omega)$ is dense into $H_0^1(\Omega)$, it suffices to prove Poincaré's inequality for functions $u \in C_c^\infty(\Omega) \subset C_c^\infty(\mathbb{R}^n)$. The fact that Ω is bounded, yields two real numbers a and b such that

$$\Omega \subset \{x = (x', x_n) \in \mathbb{R}^{n-1} \times \mathbb{R} \mid a \le x_n \le b\}.$$

We have
$$u(x', x_n) = \int_a^{x_n} \frac{\partial u}{\partial x_n}(x', t)\, dt,$$
and an application of the Cauchy–Buniakovski–Schwarz inequality gives us

$$\begin{aligned} |u(x', x_n)|^2 &\le (x_n - a) \int_a^{x_n} \left| \frac{\partial u}{\partial x_n}(x', t)\right|^2 dt \\ &\le (x_n - a) \int_{\mathbb{R}} \left| \frac{\partial u}{\partial x_n}(x', t)\right|^2 dt. \end{aligned}$$

Then
$$\int_{\mathbb{R}^{n-1}} |u(x', t)|^2\, dx' \le (x_n - a) \int_{\mathbb{R}^n} \left| \frac{\partial u}{\partial x_n}(x)\right|^2 dx,$$
which leads to

$$\int_{\mathbb{R}^n} |u(x)|^2\, dx = \int_a^b \int_{\mathbb{R}^{n-1}} |u(x', t)|^2\, dx' \le \frac{(b-a)^2}{2} \int_{\mathbb{R}^n} \left| \frac{\partial u}{\partial x_n}(x)\right|^2 dx$$

and now the assertion of Theorem C.2.2 is clear. □

By Poincaré's inequality, the inclusion $H_0^m(\Omega) \subset H^m(\Omega)$ is strict whenever Ω is bounded. Notice that $H_0^m(\mathbb{R}^n) = H^m(\mathbb{R}^n)$, due to the possibility to approximate (via mollification) the functions in $H^m(\mathbb{R}^n)$ by functions in $C_c^\infty(\mathbb{R}^n)$.

C.3 Applications to Elliptic Boundary-Value Problems

In what follows we shall illustrate the role of the variational methods in solving some problems in partial differential equations. More advanced applications may be found in books like those by G. Duvaut and J.-L. Lions [69] and I. Ekeland and R. Temam [70].

Dirichlet Problems

Let Ω be a bounded open set in $\mathbb{R}^n$ and let $f \in C(\overline{\Omega})$. A function $u \in C^2(\Omega) \cap C(\overline{\Omega})$ is said to be a *classical solution* of the Dirichlet problem

$$\begin{cases} -\Delta u + u = f & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega, \end{cases} \tag{C.2}$$

provided that it satisfies the equation and the boundary condition pointwise.

If u is a classical solution to this problem then the equation $-\Delta u + u = f$ is equivalent to

$$\int_\Omega (-\Delta u + u) \cdot v\, dx = \int_\Omega f \cdot v\, dx \quad \text{for all } v \in H_0^1(\Omega).$$

By Green's formula,

$$\int_\Omega (-\Delta u + u) \cdot v\, dx = -\int_{\partial\Omega} \frac{\partial u}{\partial n} \cdot v\, dx + \int_\Omega u \cdot v\, dx + \sum_{k=1}^n \int_\Omega \frac{\partial u}{\partial x_k} \cdot \frac{\partial v}{\partial x_k}\, dx,$$

so that we arrive at the following restatement of (C.2):

$$\sum_{k=1}^n \int_\Omega \frac{\partial u}{\partial x_k} \cdot \frac{\partial v}{\partial x_k}\, dx + \int_\Omega u \cdot v\, dx = \int_\Omega f \cdot v\, dx \tag{C.3}$$

for all $v \in C_c^\infty(\Omega)$. It turns out that (C.3) makes sense for $u \in H_0^1(\Omega)$ and $f \in L^2(\Omega)$. We shall say that a function $u \in H_0^1(\Omega)$ is a *weak solution* for the Dirichlet problem (C.2) with $f \in L^2(\Omega)$ if it satisfies (C.3) for all $v \in H_0^1(\Omega)$.

The existence and uniqueness of the weak solution for the Dirichlet problem (C.2) follows from Theorem C.1.2, applied to the functional

$$J(u) = \frac{1}{2}\, \|u\|_{H_0^1}^2 - \langle f, u\rangle_{L^2}, \quad u \in H_0^1(\Omega).$$

In fact, this functional is strictly convex and twice Gâteaux differentiable, with

$$J'(u\,; v) = \langle u, v\rangle_{H_0^1} - \langle f, v\rangle_{L^2}$$
$$J''(u\,; v, w) = \langle w, v\rangle_{H_0^1}.$$

According to Theorem C.1.2, the unique point of global minimum of J is the unique solution of the equation

$$J'(u; v) = 0 \quad \text{for all } v \in H_0^1(\Omega),$$

and clearly, the latter is equivalent with (C.3).

Neumann Problems

Let Ω be a bounded open set in $\mathbb{R}^n$ (with Lipschitz boundary) and let $f \in C(\overline{\Omega})$. A function $u \in C^2(\Omega) \cap C^1(\overline{\Omega})$ is said to be a *classical solution* of the Neumann problem

$$\begin{cases} -\Delta u + u = f & \text{in } \Omega \\ \dfrac{\partial u}{\partial n} = 0 & \text{on } \partial\Omega, \end{cases} \tag{C.4}$$

provided that it satisfies the equation and the boundary condition pointwise.

If u is a classical solution to this problem, then the equation $-\Delta u + u = f$ is equivalent to

$$\int_\Omega (-\Delta u + u) \cdot v \, dx = \int_\Omega f \cdot v \, dx \quad \text{for all } v \in H^1(\Omega),$$

and thus with

$$\sum_{k=1}^{n} \int_\Omega \frac{\partial u}{\partial x_k} \cdot \frac{\partial v}{\partial x_k} \, dx + \int_\Omega u \cdot v \, dx = \int_\Omega f \cdot v \, dx \quad \text{for all } v \in H^1(\Omega), \tag{C.5}$$

taking into account Green's formula and the boundary condition $\frac{\partial u}{\partial n} = 0$ on $\partial\Omega$. As in the case of Dirichlet problem, we can introduce a concept of a weak solution for the Neumann problem (C.4) with $f \in L^2(\Omega)$. We shall say that a function $u \in H^1(\Omega)$ is a *weak solution* for the problem (C.4) if it satisfies (C.5) for all $v \in H^1(\Omega)$.

The existence and uniqueness of the weak solution for the Neumann problem follows from Theorem C.1.2, applied to the functional

$$J(u) = \frac{1}{2} \|u\|_{H^1}^2 - \langle f, u \rangle_{L^2}, \quad u \in H^1(\Omega).$$

The details are similar to the above case of Dirichlet problem.

Corollary C.1.3 (and its generalization to finite families of functions g) allow us to prove the existence and uniqueness of considerably more subtle Neumann problems such as

$$\begin{cases} -\Delta u + u + u^3 = f & \text{in } \Omega \\ \dfrac{\partial u}{\partial n} = 0 & \text{on } \partial\Omega, \end{cases} \tag{C.6}$$

where $f \in L^2(\Omega)$. This corresponds to the case where

$$\begin{gathered} g_1(t) = \cdots = g_n(t) = t^2/2, \quad g_{n+1}(t) = t^4/4, \\ A_k u = \partial u / \partial x_k \text{ for } k = 1, \ldots, n, \quad A_{n+1} u = u, \\ p_1 = \cdots = p_n = 2, \; p_{n+1} = 4, \end{gathered}$$

and

$$J(u) = \frac{1}{2} \|u\|_{H^1}^2 + \frac{1}{4} \|u\|_{L^4}^4 - \langle f, u \rangle_{L^2}, \quad u \in V = H^1(\Omega) \cap L^4(\Omega).$$

According to Corollary C.1.3, there is a unique global minimum of J and this is done by the equation

$$J'(u\,; v) = 0 \quad \text{for all } v \in V,$$

that is, by

$$\sum_{k=1}^{n}\int_{\Omega}\frac{\partial u}{\partial x_k}\cdot\frac{\partial v}{\partial x_k}\,dx+\int_{\Omega}u\cdot v\,dx+\int_{\Omega}u^3\cdot v\,dx=\int_{\Omega}f\cdot v\,dx$$

for all $v \in V$. Notice that the latter equation represents the weak form of (C.6).

The conditions under which weak solutions provide classical solutions are discussed in textbooks like that by M. Renardy and R. C. Rogers [211].

C.4 The Galerkin Method

It is important to give here an idea how the global minimum of convex functionals can be determined via numerical algorithms. For this, consider a reflexive real Banach space V, with Schauder basis $(e_k)_k$. This means that every $u \in V$ admits a unique representation

$$u=\sum_{k=1}^{\infty}c_k e_k$$

with $c_k \in \mathbb{R}$, the convergence being in the norm topology. As a consequence, for each $n \in \mathbb{N}$ there is a linear projection

$$P_n\colon V \to V, \quad P_n u=\sum_{k=1}^{n}c_k e_k.$$

Since $P_n u \to u$ for every u, the Banach–Steinhaus theorem in functional analysis assures that $\sup \|P_n\| < \infty$.

Consider a functional $J\colon V \to \mathbb{R}$ which is twice Gâteaux differentiable and for each $u \in V$ there exist $\nabla J(u) \in V'$ and $H(u) \in L(V, V')$ such that

$$J'(u\,;v)=\langle \nabla J(u), v\rangle$$
$$J''(u\,;v,w)=\langle H(u)v, w\rangle$$

for all $u, v, w \in V$. In addition, we assume that $H(u)$ satisfies estimates of the form:

$$\begin{cases}|\langle H(u)v,w\rangle| \le M\|v\|\,\|w\|\\ \langle H(u)v,v\rangle \ge \alpha\|v\|^2\end{cases} \tag{C.7}$$

for all $u, v, w \in V$. Here M and α are positive constants.

By Taylor's formula, J is strictly convex and $\lim_{\|u\|\to\infty} J(u)=\infty$. Then, by Theorem C.1.2, J is lower semicontinuous and admits a unique global minimum.

In the *Galerkin method*, the global minimum u of J is found by a finite dimensional approximation process. More precisely, one considers the restriction of J to $V_n=\mathrm{Span}\{e_1,\dots,e_n\}$ and one computes the global minimum u_n of this restriction by solving the equation

$$\langle \nabla J(u_n), v\rangle = 0 \quad \text{for all } v \in V_n.$$

The existence of u_n follows again from Theorem C.1.2. Remarkably, these minimum points approximate the global minimum u in the following strong way:

Theorem C.4.1 *We have*

$$\lim_{n\to\infty} \|u_n - u\| = 0.$$

Proof. Letting $v_n = P_n u$, we know that $v_n \to u$. By Taylor's formula, for each n there is a $\lambda_n \in (0,1)$ such that

$$J(v_n) = J(u) + \langle \nabla J(u), v_n - u\rangle + \frac{1}{2}\langle H(u + \lambda_n(v_n - u))(v_n - u), v_n - u\rangle.$$

Combining this with the first estimate in (C.7), we get $J(v_n) \to J(u)$. By the choice of u_n, it yields that

$$J(u) \le J(u_n) \le J(v_n),$$

so that $J(u_n) \to J(u)$ too. Also, $\sup J(u_n) < \infty$.

Since $\lim_{\|u\|\to\infty} J(u) = \infty$, we deduce that the sequence $(u_n)_n$ is norm bounded. According to Theorem A.1.6, it follows that $(u_n)_n$ has a weak converging subsequence, say $u_{k(n)} \overset{w}{\to} u'$. Since J is lower semicontinuous, we have

$$J(u') \le \liminf_{n\to\infty} J(u_{k(n)}) \le J(u),$$

from which it follows that $u' = u$ and $u_n \overset{w}{\to} u$. Again by Taylor's formula, for each n there is a $\mu_n \in (0,1)$ such that

$$J(u_n) = J(u) + \langle \nabla J(u), u_n - u\rangle + \frac{1}{2}\langle H(u + \mu_n(u_n - u))(u_n - u), u_n - u\rangle.$$

This relation, when combined with the second estimate in (C.7), leads to

$$\frac{2}{\alpha}\|u_n - u\|^2 \le |J(u_n) - J(u)| + |\langle \nabla J(u), u_n - u\rangle|$$

and the conclusion of the theorem is now obvious. $\square$

D

Horn's Conjecture

In this appendix we shall deal with a problem posed by H. Weyl [244] in 1912: *Let A, B and C be Hermitian $n \times n$ matrices and denote the string of eigenvalues of A by α, where*

$$\alpha : \quad \alpha_1 \geq \cdots \geq \alpha_n,$$

and similarly write β and γ for the spectra of B and C. What α, β and γ can be the eigenvalues of the Hermitian matrices A, B and C when $C = A + B$?

There is one obvious condition, namely that the trace of C is the sum of the traces of A and B:

$$\sum_{k=1}^{n} \gamma_k = \sum_{k=1}^{n} \alpha_k + \sum_{k=1}^{n} \beta_k. \tag{D.1}$$

Weyl was able to indicate supplementary additional conditions in terms of linear inequalities on the possible eigenvalues. They will be presented in Section D.1.

Weyl's problem was studied extensively by A. Horn [111] who solved it for small n and proposed a complete set of necessary inequalities to accompany (D.1) for $n \geq 5$. Horn's inequalities have the form

$$\sum_{k \in K} \gamma_k \leq \sum_{i \in I} \alpha_i + \sum_{j \in J} \beta_j, \tag{D.2}$$

where

$$I = \{i_1, \ldots, i_r\}, \quad J = \{j_1, \ldots, j_r\}, \quad K = \{k_1, \ldots, k_r\}$$

are subsets of $\{1, \ldots, n\}$ with the same cardinality $r \in \{1, \ldots, n-1\}$ in a certain finite set T_r^n. Let us call such triplets (I, J, K) *admissible*. When $r = 1$, the condition of admissibility is

$$i_1 + j_1 = k_1 + 1.$$

If $r > 1$, this condition is:

$$\sum_{i\in I} i + \sum_{j\in J} j = \sum_{k\in K} k + \binom{r+1}{2}$$

and, for all $1 \le p \le r-1$ and all $(U,V,W) \in T_p^r$,

$$\sum_{u\in U} i_u + \sum_{v\in V} j_v = \sum_{w\in W} k_w + \binom{p+1}{2}.$$

Notice that Horn's inequalities are defined by an inductive procedure.

Horn's Conjecture *A triplet (α,β,γ) of elements of $\mathbb{R}^n_{\geq}$ occurs as eigenvalues of Hermitian matrices $A, B, C \in \mathrm{M}_n(\mathbb{C})$, with $C = A + B$, if and only if the equality* (D.1) *and Horn's inequalities* (D.2) *hold for every (I,J,K) in T_r^n, and every $r < n$.*

Nowadays this conjecture is a theorem due to recent work by A. A. Klyachko [129] and A. Knutson and T. Tao [131]. It appeals to advanced facts from algebraic geometry and representation theory (beyond the goal of this book).

The interested reader may consult the papers by R. Bhatia [27] and W. Fulton [84] for a thorough introduction to the mathematical world of Horn's conjecture.

We shall restrict here to some superficial aspects, based on the extremal property of eigenvalues.

D.1 Weyl's Inequalities

The spectrum of every Hermitian matrix $A \in \mathrm{M}_n(\mathbb{C})$ consists of n real eigenvalues $\lambda_1(A), \ldots, \lambda_n(A)$, (each counted with its multiplicity). As we shall prove later, the eigenvalues $\lambda_k(A)$ depend continuously on A. See Theorem D.1.4. The downwards/upwards rearrangements of these eigenvalues will be denoted by $\lambda_k^{\downarrow}(A)$ and $\lambda_k^{\uparrow}(A)$. Most of the time, the string of eigenvalues of A will be denoted $\alpha_1 \geq \cdots \geq \alpha_n$.

The *spectral representation theorem*, asserts that every Hermitian matrix $A \in \mathrm{M}_n(\mathbb{C})$ diagonalizes in a suitable orthonormal basis of $\mathbb{C}^n$. In fact,

$$A = \sum_{k=1}^{n} \alpha_k \langle \cdot, u_k \rangle u_k \tag{D.3}$$

for an orthonormal basis $(u_k)_k$, constituted by eigenvectors of A.

As an easy consequence, we infer the equalities

$$\alpha_1 = \max_{\|x\|=1} \langle Ax, x\rangle \tag{D.4}$$

$$\alpha_n = \min_{\|x\|=1} \langle Ax, x\rangle, \tag{D.5}$$

which in turn yield

$$\{\langle Ax, x\rangle \mid x \in \mathbb{C}^n, \|x\| = 1\} = [\alpha_n, \alpha_1]. \tag{D.6}$$

Notice that the function $x \to \langle Ax, x\rangle$ is continuous and the unit sphere is compact and connected.

The relations (D.4) and (D.5) provide the following two inequalities in Horn's list of necessary conditions:

$$\gamma_1 \le \alpha_1 + \beta_1 \tag{D.7}$$

$$\gamma_n \ge \alpha_n + \beta_n. \tag{D.8}$$

The first inequality shows that $\lambda_1^{\downarrow}(A)$ is a convex function of A, while the second shows that $\lambda_n^{\downarrow}(A)$ is concave. The two conclusions are equivalent, since

$$\lambda_k^{\downarrow}(-A) = -\lambda_{n-k+1}^{\downarrow}(A) = -\lambda_k^{\uparrow}(A). \tag{D.9}$$

A refinement of (D.4) and (D.5) is as follows:

Theorem D.1.1 (Fischer's minimax principle) *If $A \in \mathrm{M}_n(\mathbb{C})$ is a Hermitian matrix, then its eigenvalues $\alpha_1 \ge \cdots \ge \alpha_n$ can be computed by the formulae*

$$\alpha_k = \max_{\substack{V \subset \mathbb{C}^n \\ \dim V = k}} \min_{\substack{x \in V \\ \|x\|=1}} \langle Ax, x\rangle = \min_{\substack{V \subset \mathbb{C}^n \\ \dim V = n-k+1}} \max_{\substack{x \in V \\ \|x\|=1}} \langle Ax, x\rangle.$$

Proof. Let $u_1, \ldots, u_n$ be the orthonormal basis which appears in the spectral representation (D.3) of A. The vector space $W = \mathrm{Span}\{u_k, u_{k+1}, \ldots, u_n\}$ is $(n-k+1)$-dimensional and thus every k-dimensional vector subspace $V \subset \mathbb{C}^n$ will contain a point $z \in W \cap V$ with $\|z\| = 1$. According to (D.6),

$$\langle Az, z\rangle \in [\alpha_n, \alpha_k]$$

from which it follows that

$$\min_{\substack{x \in V \\ \|x\|=1}} \langle Ax, x\rangle \le \alpha_k.$$

Finally we note that equality occurs for $V = \mathrm{Span}\{u_1, \ldots, u_k\}$. $\square$

Theorem D.1.1 yields *Weyl's monotonicity principle*:

$$A \le B \text{ implies } \lambda_k^{\downarrow}(A) \le \lambda_k^{\downarrow}(B).$$

Theorem D.1.2 (Weyl's inequalities) *We have*

$$\begin{aligned} \gamma_{i+j-1} &\leq \alpha_i + \beta_j \quad \textit{if } i+j-1 \leq n \\ \gamma_{i+j-n} &\geq \alpha_i + \beta_j \quad \textit{if } i+j-n \geq 1. \end{aligned} \tag{D.10}$$

Proof. Suppose that A, B, C have the spectral representations:

$$A = \sum_{k=1}^{n} \alpha_k \langle \cdot , u_k \rangle u_k; \quad B = \sum_{k=1}^{n} \beta_k \langle \cdot , v_k \rangle v_k; \quad C = \sum_{k=1}^{n} \gamma_k \langle \cdot , w_k \rangle w_k.$$

Since

$$\dim \operatorname{Span}\{u_i, \ldots, u_n\} + \dim \operatorname{Span}\{v_j, \ldots, v_n\} + \dim \operatorname{Span}\{w_1, \ldots, w_{i+j-1}\}$$

is $(n-i+1)+(n-j+1)+(i+j-1) = 2n+1$, the above three spaces must have in common a vector x with $\|x\| = 1$. Then, according to (D.4) and (D.5),

$$\langle Ax, x \rangle \leq \alpha_i, \quad \langle Bx, x \rangle \leq \beta_j, \quad \langle (A+B)x, x \rangle \geq \gamma_{i+j-1}$$

and the proof is complete. □

Corollary D.1.3 *The following inequalities hold:*

$$\alpha_i + \beta_n \leq \gamma_i \leq \alpha_i + \beta_1.$$

Theorem D.1.4 (Weyl's perturbation theorem) *For every pair of Hermitian matrices A, B in $\mathrm{M}_n(\mathbb{R})$, we have*

$$\max_{1 \leq k \leq n} |\lambda_k(A) - \lambda_k(B)| \leq \|A - B\|.$$

Particularly, the eigenvalues $\lambda_k(A)$ are continuous functions of A.

Proof. In fact, for every Hermitian matrix A we have

$$\|A\| = \sup_{\|x\|=1} |\langle Ax, x \rangle| = \max\{|\lambda_1^{\downarrow}(A)|, |\lambda_n^{\downarrow}(A)|\}.$$

Consequently, by applying Corollary D.1.3 to A, $B - A$ and $C = B$, we get

$$\lambda_k(A) - \|B - A\| \leq \lambda_k(B) \leq \lambda_k(A) + \|B - A\|.$$

□

Weyl's perturbation theorem has applications to numerical analysis.

D.2 The Case $n = 2$

In this section we indicate a proof to Horn's conjecture in the case of 2×2 Hermitian matrices. In this case the set of Horn's inequalities reduces to Weyl's inequalities (D.10). We shall prove that for all families of real numbers $\alpha_1 \geq \alpha_2$, $\beta_1 \geq \beta_2$, $\gamma_1 \geq \gamma_2$, which verify Weyl's inequalities,

$$\gamma_1 \leq \alpha_1 + \beta_1 \quad \gamma_2 \leq \alpha_2 + \beta_1 \quad \gamma_2 \leq \alpha_1 + \beta_2,$$

and the trace formula (D.1),

$$\gamma_1 + \gamma_2 = \alpha_1 + \alpha_2 + \beta_1 + \beta_2,$$

there exist symmetric matrices $A, B, C \in \mathrm{M}_2(\mathbb{R})$ with $C = A + B$, $\sigma(A) = (\alpha_1, \alpha_2)$, $\sigma(B) = (\beta_1, \beta_2)$ and $\sigma(C) = (\gamma_1, \gamma_2)$.

Assume, for the sake of simplicity, that the spectra of A and B are respectively $\alpha = (4, 2)$ and $\beta = (2, -2)$. Then the conditions above may be read as

$$\gamma_1 + \gamma_2 = 6, \quad \gamma_1 \geq \gamma_2 \tag{D.11}$$

$$\gamma_1 \leq 6, \quad \gamma_2 \leq 2. \tag{D.12}$$

This shows that γ has the form $\gamma = (6-a, a)$, with $0 \leq a \leq 2$; clearly, $\gamma_1 \geq \gamma_2$. We shall prove that every pair $(6 - a, a)$ with $0 \leq a \leq 2$ can be the spectrum of a sum $A + B$.

In fact, the relations (D.11) and (D.12) lead us to consider (in the plane $0\gamma_1\gamma_2$) the line segment XY, where $X = (6, 0)$ and $Y = (4, 2)$. Starting with the matrices

$$A = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix}$$

and

$$R_\theta^\star \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} R_\theta,$$

where

$$R_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix},$$

we should remark that the spectrum $(\lambda_1^\downarrow(C_\theta), \lambda_2^\downarrow(C_\theta))$ of the matrix

$$C_\theta = \begin{pmatrix} 4 & 0 \\ 0 & 2 \end{pmatrix} + R_\theta^\star \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} R_\theta$$

lies on the line segment XY for all $\theta \in [0, \pi/2]$. In fact, since the eigenvalues of a matrix are continuous functions on the entries of that matrix, the map

$$\theta \to (\lambda_1^\downarrow(C_\theta), \lambda_2^\downarrow(C_\theta))$$

is continuous. The trace formula shows that the image of this map is a subset of the line $\gamma_1 + \gamma_2 = 6$. X corresponds to $\theta = 0$, and Y corresponds to $\theta = \pi/2$. Since the image should be a line segment, we conclude that each point of XY represents the spectrum of a matrix C_θ with $\theta \in [0, \pi/2]$.

D.3 Majorization Inequalities and the Case $n = 3$

According to Schur's Theorem (Theorem 1.10.2), if A is Hermitian matrix with diagonal elements $a_{11}, \ldots, a_{nn}$ and eigenvalues $\lambda_1, \ldots, \lambda_n$, then

$$(a_{11}, \ldots, a_{nn}) \prec (\lambda_1, \ldots, \lambda_n).$$

Since the spectrum is invariant under unitary equivalence of matrices, this result yields a maximum principle, first observed by Ky Fan:

$$\sum_{k=1}^{r} \alpha_k = \max_{\substack{(x_k)_{k=1}^r \\ \text{orthonormal} \\ \text{family}}} \sum_{k=1}^{r} \langle Ax_k, x_k \rangle \quad \text{for } r = 1, \ldots, n.$$

Particularly, the sums $\sum_{k=1}^{r} \lambda_k^{\downarrow}(A)$ are convex functions on A. This leads to *Ky Fan's inequalities*:

$$\sum_{k=1}^{r} \gamma_k \leq \sum_{k=1}^{r} \alpha_k + \sum_{k=1}^{r} \beta_k, \quad \text{for } r = 1, \ldots, n, \tag{D.13}$$

which can be restated as

$$\lambda(A + B) \prec \lambda^{\downarrow}(A) + \lambda^{\downarrow}(B). \tag{D.14}$$

The complementary inequality,

$$\lambda^{\downarrow}(A) + \lambda^{\uparrow}(B) \prec \lambda(A + B), \tag{D.15}$$

also works and it was proved in an equivalent form by V. B. Lidskii [143] and later by H. Wielandt [246]:

Theorem D.3.1 (Lidskii–Wielandt inequalities) *Let A, B, C be three Hermitian matrices with $C = A + B$. Then for every $1 \leq r \leq n$ and every $1 \leq i_1 < \cdots < i_r \leq n$ we have the inequalities*

$$\sum_{k=1}^{r} \gamma_{i_k} \leq \sum_{k=1}^{r} \alpha_{i_k} + \sum_{k=1}^{r} \beta_k \tag{D.16}$$

as well as the corresponding inequalities obtained by interchanging A and B.

Proof. C. K. Li and R. Mathias [142] We must prove the inequality

$$\sum_{k=1}^{r} [\lambda_{i_k}^{\downarrow}(A + B) - \lambda_{i_k}^{\downarrow}(A)] \leq \sum_{k=1}^{r} \lambda_k^{\downarrow}(B). \tag{D.17}$$

Without loss of generality we may assume that $\lambda_r^{\downarrow}(B) = 0$; for this, replace B by $B - \lambda_r^{\downarrow}(B) \cdot I$.

Let $B = B^+ - B^-$ be the canonical decomposition of B into the positive and negative parts. Since $B \leq B^+$, Weyl's monotonicity principle yields $\lambda^{\downarrow}_{i_k}(A+B) \leq \lambda^{\downarrow}_{i_k}(A+B^+)$, so that the left hand side of (D.17) is

$$\sum_{k=1}^{r} [\lambda^{\downarrow}_{i_k}(A+B^+) - \lambda^{\downarrow}_{i_k}(A)],$$

which in turn is less than or equal to

$$\sum_{k=1}^{n} [\lambda^{\downarrow}_{k}(A+B^+) - \lambda^{\downarrow}_{k}(A)] = \mathrm{trace}(B^+).$$

Or, $\mathrm{trace}(B^+) = \sum_{k=1}^{r} \lambda^{\downarrow}_k(B)$ since $\lambda^{\downarrow}_r(B) = 0$. □

We are now in a position to list all the Horn inequalities in the case of 3×3-dimensional Hermitian matrices:

- *Weyl's inequalities*,

$$\begin{array}{lll} \gamma_1 \leq \alpha_1 + \beta_1 & \gamma_2 \leq \alpha_1 + \beta_2 & \gamma_2 \leq \alpha_2 + \beta_1 \\ \gamma_3 \leq \alpha_1 + \beta_3 & \gamma_3 \leq \alpha_3 + \beta_1 & \gamma_3 \leq \alpha_2 + \beta_2; \end{array}$$

- *Ky Fan's inequality*,

$$\gamma_1 + \gamma_2 \leq \alpha_1 + \alpha_2 + \beta_1 + \beta_2;$$

- *Lidskii–Wielandt inequalities* (taking into account the symmetric role of A and B),

$$\begin{aligned} \gamma_1 + \gamma_3 &\leq \alpha_1 + \alpha_3 + \beta_1 + \beta_2 \\ \gamma_2 + \gamma_3 &\leq \alpha_2 + \alpha_3 + \beta_1 + \beta_2 \\ \gamma_1 + \gamma_3 &\leq \alpha_1 + \alpha_2 + \beta_1 + \beta_3 \\ \gamma_2 + \gamma_3 &\leq \alpha_1 + \alpha_2 + \alpha_3 + \beta_2 + \beta_3; \end{aligned}$$

- *Horn's inequality*,

$$\gamma_2 + \gamma_3 \leq \alpha_1 + \alpha_3 + \beta_1 + \beta_3.$$

The last inequality follows from (D.15), which in the case $n = 3$ may be read as

$$(\alpha_1 + \beta_3, \alpha_2 + \beta_2, \alpha_3 + \beta_1) \prec (\gamma_1, \gamma_2, \gamma_3).$$

Adding to the above twelve inequalities the trace formula

$$\gamma_1 + \gamma_2 + \gamma_3 = \alpha_1 + \alpha_2 + \alpha_3 + \beta_1 + \beta_2 + \beta_3$$

we get a set of necessary and sufficient conditions for the existence of three symmetric matrices $A, B, C \in \mathrm{M}_3(\mathbb{R})$, with $C = A + B$, and spectra equal respectively to

$$\alpha_1 \geq \alpha_2 \geq \alpha_3; \quad \beta_1 \geq \beta_2 \geq \beta_3; \quad \gamma_1 \geq \gamma_2 \geq \gamma_3.$$

The proof is similar to the case $n = 2$. The eigenvalues of $A + UBU^\star$, as U varies over unitary matrices, is a convex polyhedron in $\mathbb{R}^3$ and this polyhedron is described by (D.1) and Horn's inequalities.

For larger n, things become much more intricate. For example, for $n = 7$, there are 2062 such inequalities, not all of them independent.

As the cases $n = 2$ and $n = 3$ suggest, Horn's conjecture is a problem of intersections. And indeed, the Schubert calculus in algebraic geometry proved to be at the heart of the matter. The reader is urged to read the paper by R. Bhatia [27] to get the flavor of the mathematics implied in this solution.

Needless to say, many other related problems have been solved with this occasion. The paper by W. Fulton [84] contains a good account on this matter. We end by recalling here the multiplicative companion to Horn's inequalities:

Theorem D.3.2 *Let $\alpha_1 \geq \alpha_2 \geq \alpha_3$, $\beta_1 \geq \beta_2 \geq \beta_3$, $\gamma_1 \geq \gamma_2 \geq \gamma_3$, be triplets of nonnegative real numbers. Then there exist matrices A and B with singular numbers $s_k(A) = \alpha_k$, $s_k(B) = \beta_k$, $s_k(AB) = \gamma_k$, if and only if*

$$\prod_{k \in K} \gamma_k \leq \prod_{i \in I} \alpha_i \prod_{j \in J} \beta_j$$

for all admissible triplets (I, J, K).

References

1. S. Abramovich, G. Jameson and G. Sinnamon, *Refining Jensen's Inequality*, Bull. Math. Soc. Sci. Math. Roumanie **47(95)** (2004), 3–14.
2. J. Aczél, *The notion of mean values*, Norske Vid. Selsk. Forhdl., Trondhjem **19** (1947), 83–86.
3. J. Aczél, *A Generalization of the Notion of Convex Functions*, Norske Vid. Selsk. Forhdl., Trondhjem **19** (1947), 87–90.
4. G. Alberti and L. Ambrosio, *A geometrical approach to monotone functions in* $\mathbb{R}^n$, Math. Z. **230** (1999), 259–316.
5. A. D. Alexandrov, *Almost everywhere existence of the second differential of a convex function and some properties of convex surfaces connected to it*, Leningrad State Univ. Ann., Math. Ser. **6** (1939), 3–35. (Russian)
6. H. W. Alt, *Lineare Funktionalanalysis. Eine anwendungsorientierte Einführung*, 2 Auflage, Springer-Verlag, 1992.
7. H. Alzer, *On an integral inequality*, Anal. Numér. Théor. Approx. **18** (1989), 101–103.
8. T. Ando, C.-K. Li and R. Mathias, *Geometric Means*, Linear Algebra Appl. **385** (2004), 305–334.
9. G. Andrews, R. Askey and R. Roy, *Special Functions*, Cambridge Univ. Press, 1999.
10. E. Artin, *The Gamma Function*, Holt, Rinehart and Winston, New York, 1964. English translation of German original, *Einführung in die Theorie der Gammafunktion*, Teubner, 1931.
11. E. Asplund, *Fréchet differentiability of convex functions*, Acta Math. **121** (1968), 31–47.
12. M. Atiyah, *Angular momentum, convex polyhedra and algebraic geometry*, Proc. Edinburgh Math. Soc. **26** (1983), 121–138.
13. G. Aumann, *Konvexe Funktionen und die Induktion bei Ungleichungen swischen Mittelwerten*, S.-B. math.-naturw. Abt. Bayer. Akad. Wiss. München (1933), 403–415.
14. V. Barbu and Th. Precupanu, *Convexity and Optimization in Banach Spaces*, Ed. Academiei, Bucharest, and D. Reidel Publ. Co., Dordrecht, 1986.

15. F. Barthe, *Inégalités de Brascamp–Lieb et convexité*, C. R. Acad. Sci. Paris, Sér. I Math. **324** (1997), 885–887.
16. F. Barthe, *Optimal Young's Inequality and Its Converse: A Simple Proof*, Geom. Funct. Anal. **8** (1998), 234–242.
17. S. Barza and C. P. Niculescu, *Integral Inequalities for Concave Functions,* Publ. Math. Debrecen **68** (2006), to appear.
18. E. F. Beckenbach and R. Bellman, *Inequalities*, 2nd edition, Springer-Verlag, Berlin, 1983.
19. R. Benedetti and J.-J. Risler, *Real algebraic and semi-algebraic sets*, Actualités Mathématiques, Hermann, Paris, 1990.
20. C. Bennett and R. Sharpley, *Interpolation of Operators*, Academic Press, 1988.
21. G. Berkhan, *Zur projektivischen Behandlung der Dreiecksgeometrie*, Arch. Math. Phys. (3) **11** (1907), 1–31.
22. J. Bernoulli, *Positiones Arithmeticae de seriebus infinitas, earumque summa finita*, Basileae, 1689, *Opera* 1, 375–402.
23. M. Berger, *Convexity*, Amer. Math. Monthly **97** (1990), 650–678.
24. L. Berwald, *Verallgemeinerung eines Mittelwertsatzes von J. Favard, für positive konkave Funktionen*, Acta Math. **79** (1947), 17–37.
25. M. Bessenyei and Z. Páles, *Higher order generalizations of Hadamard's inequality*, Publ. Math. Debrecen **61** (2002), 623–643.
26. M. Bessenyei and Z. Páles, *Hadamard-type inequalities for generalized convex functions*, Math. Inequal. Appl. **6** (2003), 379–392.
27. R. Bhatia, *Linear Algebra to Quantum Cohomology: The Story of Alfred Horn's Inequalities*, Amer. Math. Monthly **108** (2001), 289–318.
28. G. Bianchi, A. Colesanti and C. Pucci, *On the Second Differentiability of Convex Surfaces*, Geom. Dedicata **60** (1996), 39–48.
29. Z. W. Birnbaum and W. Orlicz, *Über die Verallgemeinerung des Begriffes der zueinander konjugierten Potenzen*, Studia Math. **3** (1931), 1–67.
30. E. Bishop and K. de Leeuw, *The representation of linear functionals by measures on sets of extreme points*, Ann. Inst. Fourier (Grenoble) **9** (1959), 305–331.
31. H. Blumberg, *On convex functions*, Trans. Amer. Math. Soc. **20** (1919), 40–44.
32. H. Bohr and J. Mollerup, *Laerebog i Mathematisk Analyse III*, Kopenhagen, 1922.
33. Ch. Borell, *Convex set functions in d-space*, Period. Math. Hungar. **6** (1975), 111–136.
34. Ch. Borell, *Diffusion Equations and Geometric Inequalities*, Potential Anal. **12** (2000), 49–71.
35. D. Borwein, J. Borwein, G. Fee and R. Girgensohn, *Refined convexity and special cases of the Blaschke–Santalo inequality*, Math. Inequal. Appl. **4** (2001), 631–638.
36. J. Borwein, *Continuity and differentiability properties of convex operators*, Proc. London Math. Soc. **44** (1982), 420–444.
37. J. M. Borwein and P. B. Borwein, *The Way of All Means*, Amer. Math. Monthly **94** (1987), 519–522.

38. J. M. Borwein and A. S. Lewis, *Convex Analysis and Nonlinear Optimization. Theory and Examples*, Springer-Verlag, Berlin, 2000.
39. H. J. Brascamp and E. H. Lieb, *Best constants in Young's inequality, its converse and its generalization to more than three functions*, Advances in Math. **20** (1976), 151–173.
40. H. J. Brascamp and E. H. Lieb, *On Extensions of the Brunn–Minkowski and Prékopa–Leindler Theorems, Including Inequalities for Log Concave Functions, and with an Application to the Diffusion Equation*, J. Functional Analysis **22** (1976), 366–389.
41. Y. Brenier, *Polar Factorization and Monotone Rearrangemenent of Vector-Valued Functions*, Comm. Pure Appl. Math. **44** (1991), 375–417.
42. J. L. Brenner and B. C. Carlson, *Homogeneous mean values: weights and asymptotics*, J. Math. Anal. Appl. **123** (1987), 265–280.
43. H. Brezis, *Opérateurs maximaux monotones et semigroups de contractions dans les espaces de Hilbert*, North Holland, Amsterdam, 1973.
44. P. S. Bullen, D. S. Mitrinović and P. M. Vasić, *Means and Their Inequalities*, D. Reidel Publishing Company, Dordrecht, 1988.
45. Y. D. Burago and V. A. Zalgaller, *Geometric Inequalities*, Springer-Verlag, New York, 1988.
46. F. Burk, *The Geometric, Logarithmic and Arithmetic Mean Inequality*, Amer. Math. Monthly **94** (1987), 527–528.
47. G. T. Cargo, *Comparable means and generalized convexity*, J. Math. Anal. Appl. **12** (1965), 387–392.
48. T. Carleman, *Sur les fonctions quasi-analytiques.* In *Comptes Rendus du 5e Congrès des Mathématiciens Scandinaves*, Helsingfors, Finland, 1922, 181–196.
49. B. C. Carlson, *Algorithms involving arithmetic and geometric means*, Amer. Math. Monthly **78** (1971), 496–505.
50. A.-L. Cauchy, *Cours d'analyse de l'Ecole Royale Polytechnique*, 1ère partie, *Analyse algébrique*, Paris, 1821. See also *Œuvres complètes*, IIe série, VII.
51. J. Céa, *Optimisation. Théorie et Algorithmes*, Dunod, Paris, 1971.
52. S. H. Chang, *On the distribution of the characteristic values and singular values of linear integral equations*, Trans. Amer. Math. Soc. **67** (1949), 351–67.
53. G. Choquet, *Les cônes convexes faiblement complets dans l'analyse*, Proc. Intern. Congr. Mathematicians, Stockholm, 1962, 317–330.
54. A. Čižmešija and J. E. Pečarić, *Mixed means and Hardy's inequality*, Math. Inequal. Appl. **1** (1998), 491–506.
55. A. Čižmešija, J. E. Pečarić and L.-E. Persson, *On strengthened Hardy and Pólya–Knopp's inequalities*, J. Approx. Theory **125** (2003), 74–84.
56. F. H. Clarke, *Optimization and Nonsmooth Analysis*, 2nd edition, Classics in Applied Mathematics, Vol. 5, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990.
57. J. A. Clarkson, *Uniformly convex spaces*, Trans. Amer. Math. Soc. **40** (1936), 396–414.

58. J. B. Conway, *A Course in Functional Analysis*, 2nd edition, Springer-Verlag, Berlin, 1997.
59. D. Cordero-Erausquin, *Some Applications of Mass Transport to Gaussian-Type Inequalities*, Arch. Rational Mech. Anal. **161** (2002), 257–269.
60. D. Cordero-Erausquin, R. J. McCann and M. Schmuckenschläger, *A Riemannian interpolation inequality à la Borell, Brascamp and Lieb*, Invent. Math. **146** (2001), 219–257.
61. M. G. Crandall, H. Ishii and P.-L. Lions, *User's guide to viscosity solutions of second order partial differential equations*, Bull. Amer. Math. Soc. **27** (1992), 1–67.
62. S. Dancs and B. Uhrin, *On a class of integral inequalities and their measure-theoretic consequences*, J. Math. Anal. Appl. **74** (1980), 388–400.
63. K. R. Davidson and A. P. Donsig, *Real Analysis with Real Applications*, Prentice Hall, Upper Saddle River, N.J., 2002.
64. M. M. Day, *Normed linear spaces*, 3rd edition, Springer-Verlag, Berlin, 1973.
65. J. B. Diaz and F. T. Metcalf, *Stronger forms of a class of inequalities of G. Pólya–G. Szegö, and L. V. Kantorovich*, Bull. Amer. Math. Soc. **69** (1963), 415–418.
66. J. Dieudonné, *Foundations of Modern Analysis*, Academic Press, 1960.
67. S. S. Dragomir and N. M. Ionescu, *Some converse of Jensen's inequality and applications*, Anal. Numér. Théor. Approx. **23** (1994), 71–78.
68. J. Duncan and C. M. McGregor, *Carleman's inequality*, Amer. Math. Monthly **110** (2003), 424–431.
69. G. Duvaut and J.-L. Lions, *Inequalities in Mechanics and Physics*, Springer-Verlag, Berlin, 1976.
70. I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*, North-Holland Publishing Company, Amsterdam, 1976.
71. N. Elezović and J. Pečarić, *Differential and integral F-means and applications to digamma function*, Math. Inequal. Appl. **3** (2000), 189–196.
72. A. Engel, *Problem Solving Strategies*, Springer-Verlag, Berlin, 1998.
73. Euclid, *The thirteen books of Euclid's Elements* (translated by Sir Thomas Heath, Cambridge, 1908).
74. L. C. Evans and R. F. Gariepy, *Measure Theory and Fine Properties of Functions*, CRC Press, Boca Raton–New York–London–Tokyo, 1992.
75. K. Fan, *On a theorem of Weyl concerning eigenvalues of linear transformations II*, Proc. Nat. Acad. Sci. USA **36** (1950), 31–35.
76. J. Favard, *Sur les valeurs moyennes*, Bull. Sci. Math. **57** (1933), 54–64.
77. H. Federer, *Geometric Measure Theory*, Springer-Verlag, New York, 1969.
78. W. Fenchel, *On conjugate convex functions*, Canad. J. Math. **1** (1949), 73–77.

79. W. Fenchel, *Convex Cones, Sets and Functions* (mimeographed lecture notes), Princeton Univ. Press, Princeton, 1951.
80. B. de Finetti, *Sul concetto di media*, Giornale dell' Instituto Italiano degli Attuari **2** (1931), 369–396.
81. A. M. Fink, *A best possible Hadamard inequality*, Math. Inequal. Appl. **1** (1998), 223–230.
82. J. Franklin, *Mathematical Methods of Economics*, Amer. Math. Monthly **90** (1983), 229–244.
83. L. Fuchs, *A new proof of an inequality of Hardy, Littlewood and Pólya*, Mat. Tidsskr. B., 1947, 53–54.
84. W. Fulton, *Eigenvalues, invariant factors, highest weights and Schubert calculus*, Bull. Amer. Math. Soc. **37** (2000), 209–249.
85. D. Gale, V. Klee and R. T. Rockafellar, *Convex functions on convex polytopes*, Proc. Amer. Math. Soc. **19** (1968), 867–873.
86. L. Galvani, *Sulle funzioni converse di una o due variabili definite in aggregate qualunque*, Rend. Circ. Mat. Palermo **41** (1916), 103–134.
87. R. J. Gardner, *The Brunn–Minkowski inequality: A survey with proofs*. Preprint, 2001; available at http://www.ac.wwu.edu/~gardner/research.html
88. R. J. Gardner, *The Brunn–Minkowski inequality*, Bull. Amer. Math. Soc. **39** (2002), 355–405.
89. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, Academic Press, 1996. (CD-ROM version)
90. D. Gronau and J. Matkowski, *Geometrical convexity and generalizations of the Bohr–Mollerup theorem on the gamma function*, Math. Pannon. **4** (1993), 153–160.
91. D. Gronau and J. Matkowski, *Geometrically convex solutions of certain difference equations and generalized Bohr–Mollerup type theorems*, Results Math. **26** (1994), 290–297.
92. P. M. Gruber, *Aspects of convexity and its applications*, Expo. Math. **2** (1984), 47–83.
93. P. M. Gruber and J. M. Willis (eds.), *Handbook of convex geometry*, North-Holland, Amsterdam, 1993.
94. J. Hadamard, *Étude sur les propriétés des fonctions entières et en particulier d'une fonction considerée par Riemann*, J. Math. Pures Appl. **58** (1893), 171–215.
95. G. Hamel, *Eine Basis aller Zahlen und die unstetigen Lösungen der Funktionalgleichung* $f(x+y) = f(x) + f(y)$, Math. Ann. **60** (1905), 459–462.
96. O. Hanner, *On the uniform convexity of* L^p *and* l^p, Ark. Mat. **3** (1955), 239–244.
97. G. H. Hardy, *A note on two inequalities*, J. London Math. Soc. **11** (1936), 167–170.
98. G. H. Hardy, J. E. Littlewood and G. Pólya, *Some simple inequalities satisfied by convex functions*, Messenger Math. **58** (1929), 145–152.
99. G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, 2nd edition, Cambridge Mathematical Library, 1952; reprinted 1988.
100. H. Heinig and L. Maligranda, *Weighted inequalities for monotone and concave functions*, Studia Math. **116** (1995), 133–165.

101. R. Henstock and A. M. Macbeath, *On the measure of sum sets I. The theorems of Brunn, Minkowski and Lusternik*, Proc. London Math. Soc. **3** (1953), 182–194.
102. Ch. Hermite, *Sur deux limites d'une intégrale définie*, Mathesis **3** (1883), 82.
103. E. Hewitt and K. Stromberg, *Real and Abstract Analysis*, Springer-Verlag, Berlin, 1965.
104. J.-B. Hiriart-Urruty, *Ensembles de Tchebychev vs. ensembles convexes: l'état de la situation vu via l'analyse non lisse*, Ann. Sci. Math. Québec **22** (1998), 47–62.
105. J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, Springer-Verlag, Berlin, 1993.
106. O. Hölder, *Über einen Mittelwertsatz*, Nachr. Ges. Wiss. Goettingen, 1889, 38–47.
107. L. Hörmander, *Sur la fonction d'appui des ensembles convexes dans une espace localement convexe*, Ark. Mat. **3** (1954), 181–186.
108. L. Hörmander, *Notions of Convexity*, Birkhäuser, Boston, 1994.
109. A. Horn, *Doubly stochastic matrices and the diagonal of a rotation matrix*, Amer. J. Math. **76** (1954), 620–630.
110. A. Horn, *On the eigenvalues of a matrix with prescribed singular values*, Proc. Amer. Math. Soc. **5** (1954), 4–7.
111. A. Horn, *Eigenvalues of sums of Hermitian matrices*, Pacific J. Math. **12** (1962), 225–241.
112. H. Hornich, *Eine Ungleichung für Vektorlängen*, Math. Z. **48** (1942), 268–273.
113. K. S. K. Iyengar, *Note on an inequality*, Math. Student **6** (1938), 75–76.
114. J. L. W. V. Jensen, *Om konvexe Funktioner og Uligheder mellem Middelvaerdier*, Nyt. Tidsskr. Math. **16B** (1905), 49–69.
115. J. L. W. V. Jensen, *Sur les fonctions convexes et les inegalités entre les valeurs moyennes*, Acta Math. **30** (1906), 175–193.
116. M. Johansson, L.-E. Persson and A. Wedestig, *Carleman's inequality-history, proofs and some new generalizations*, J. Inequal. Pure Appl. Math. (JIPAM) **4** (2003), no. 3, article 53.
117. B. Josefson, *Weak sequential convergence in the dual of a Banach space does not imply norm convergence*, Ark. Mat. **13** (1975), 79–89.
118. S. Kaijser, L.-E. Persson and A. Öberg, *On Carleman and Knopp's inequalities*, J. Approx. Theory **117** (2002), 140–151.
119. J. Karamata, *Sur une inégalité relative aux fonctions convexes*, Publ. Math. Univ. Belgrade **1** (1932), 145–148.
120. N. D. Kazarinoff, *Analytic Inequalities*, Holt, Rinehart and Winston, New York, 1961.
121. B. Kawohl, *When are superharmonic functions concave? Applications to the St. Venant torsion problem and to the fundamental mode of the clamped membrane*, Z. Angew. Math. Mech. **64** (1984), 364–366.
122. B. Kawohl, *Rearrangements and Convexity of Level Sets in PDE*, Lecture Notes in Math. **1150**, Springer-Verlag, 1985.
123. K. Kedlaya, *Proof of a Mixed Arithmetic-Mean, Geometric-Mean Inequality*, Amer. Math. Monthly **101** (1994), 355–357.

124. J. L. Kelley, *General Topology*, D. van Nostrand Company, Princeton, 1957.
125. L. G. Khanin, *Problem M 1083*, Kvant **18** (1988), no. 1, 35; Kvant **18** (1988), no. 5, 35.
126. C. H. Kimberling, *Some Corollaries to an Integral Inequality*, Amer. Math. Monthly **81** (1974), 269–270.
127. M. S. Klamkin, *Inequalities for inscribed and circumscribed polygons*, Amer. Math. Monthly **87** (1980), 469–473.
128. V. Klee, *Some new results on smoothness and rotundity in normed linear spaces*, Math. Ann. **139** (1959), 51–63.
129. A. A. Klyachko, *Stable bundles, representation theory and Hermitian operators*, Selecta Math. **4** (1998), 419–445.
130. K. Knopp, *Über Reihen mit positiven Gliedern*, J. London Math. Soc. **3** (1928), 205–211.
131. A. Knutson and T. Tao, *The honeycomb model of* $\mathrm{GL}_n(\mathbb{C})$ *tensor products I. Proof of the saturation conjecture*, J. Amer. Math. Soc. **12** (1999), 1055–1090.
132. M. A. Krasnosel'skii and Ya. B. Rutickii, *Convex Functions and Orlicz Spaces*, P. Nordhoff, Groningen, 1961.
133. J. L. Krivine, *Théorèmes de factorisation dans les espaces reticulés*, Séminaire Maurey–Schwartz, 1973–74, Exposés 22–23, Ecole Polytechnique, Paris.
134. F. Kubo and T. Ando, *Means of Positive Linear Operators*, Math. Ann. **246** (1980), 205–224.
135. A. Kufner and L.-E. Persson, *Weighted Inequalities of Hardy Type*, World Scientific, New Jersey–London–Singapore–Hong Kong, 2003.
136. J. Lamperti, *On the isometries of certain function-spaces*, Pacific J. Math. **8** (1958), 459–466.
137. S. Lang, *Analysis I*, Addison–Wesley Publ. Co., Reading, Massachusetts, 1968.
138. S. Lang, *Analysis II*, Addison–Wesley Publ. Co., Reading, Massachusetts, 1969.
139. J. M. Lasry and P.-L. Lions, *A remark on regularization in Hilbert spaces*, Israel J. Math. **55** (1986), 257–266.
140. J. D. Lawson and Y. Lim, *The Geometric Mean, Matrices, Metrics and More*, Amer. Math. Monthly **108** (2001), 797–812.
141. V. I. Levin and S. B. Stečkin, *Inequalities*, Amer. Math. Soc. Transl. **14** (1960), 1–22.
142. C.-K. Li and R. Mathias, *The Lidskii–Mirsky–Wielandt theorem—additive and multiplicative versions*, Numer. Math. **81** (1999), 377–413.
143. V. B. Lidskii, *The proper values of the sum and product of symmetric matrices*, Dokl. Akad. Nauk SSSR **74** (1950), 769–772.
144. E. H. Lieb and M. Loss, *Analysis*, 2nd edition, Amer. Math. Soc., Providence, R.I., 2001.
145. T.-P. Lin, *The power mean and the logarithmic mean*, Amer. Math. Monthly **81** (1974), 879–883.
146. J. Lindenstrauss and A. Pełczyński, *Absolutely summing operators in* $\mathcal{L}_p$*-spaces and their applications*, Studia Math. **29** (1968), 275–326.

147. J. Lindenstrauss and L. Tzafriri, *Classical Banach spaces*, Vol. 1 (1977); Vol. 2 (1979), Springer-Verlag, Berlin.
148. A. Lyapunov, *Nouvelle forme du théorème sur la limite de probabilité*, Mémoires de l'Acad. de St.-Petersburg (VIII) **12** (1901), no. 5, 1–24.
149. C. Maclaurin, *A second letter to Martin Folkes, Esq.; concerning the roots of equations, with the demonstration of other rules in algebra*, Philos. Transactions **36** (1729), 59–96.
150. S. M. Malamud, *Some complements to the Jensen and Chebyshev inequality and a problem of W. Walter*, Proc. Amer. Math. Soc. **129** (2001), 2671–2678.
151. L. Maligranda, *Concavity and convexity of rearrangements*, Comment. Math. Prace Mat. **32** (1992), 85–90.
152. L. Maligranda, *Why Hölder's inequality should be called Rogers' inequality*, Math. Inequal. Appl. **1** (1998), 69–83.
153. L. Maligranda, J. E. Pečarić and L.-E. Persson, *On Some Inequalities of the Grüss–Barnes and Borell Type*, J. Math. Anal. Appl. **187** (1994), 306–323.
154. M. Marcus and J. Lopes, *Inequalities for symmetric functions and Hermitian matrices*, Canad. J. Math. **8** (1956), 524–531.
155. A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*, Academic Press, 1979.
156. M. Matić and J. E. Pečarić, *Some companion inequalities to Jensen's inequality*, Math. Inequal. Appl. **3** (2000), 355–368.
157. J. Matkowski, *Convex and affine functions with respect to a mean and a characterization of the weighted quasi-arithmetic means*, Real Anal. Exchange **29** (2003/04), 229–246.
158. J. Matkowski and J. Rätz, *Convexity of power functions with respect to symmetric homogeneous means*, Internat. Ser. Numer. Math. **123** (1997), 231–247.
159. J. Matkowski and J. Rätz, *Convex functions with respect to an arbitrary mean*, Internat. Ser. Numer. Math. **123** (1997), 249–258.
160. S. Mazur and S. Ulam, *Sur les transformations isométriques d'espaces vectoriels normés*, C. R. Acad. Sci. Paris **194** (1932), 946–948.
161. R. J. McCann, *Existence and uniqueness of monotone measure-preserving maps*, Duke Math. J. **80** (1995), 309–323.
162. R. J. McCann, *A convexity principle for interacting gases*, Adv. Math. **128** (1997), 153–179.
163. F. Mignot, *Contrôle dans les Inéquations Variationelles Elliptiques*, J. Funct. Anal. **22** (1976), 130–185.
164. H. Minkowski, *Theorie der Konvexen Körper, Insbesondere Begründung ihres Oberflächenbegriffs*, Gesammelte Abhandlungen II, Leipzig, 1911.
165. G. Minty, *Monotone nonlinear operators on a Hilbert space*, Duke Math. J. **29** (1962), 341–346.
166. D. S. Mitrinović, *Analytic Inequalities*, Springer-Verlag, Berlin and New York, 1970.
167. D. S. Mitrinović and I. B. Lacković, *Hermite and convexity*, Aequationes Math. **28** (1985), 229–232.

168. D. S. Mitrinović, J. E. Pečarić and A. M. Fink, *Inequalities Involving Functions and Their Integrals and Derivatives*, Kluwer Academic Publ., Dordrecht, 1991.
169. D. S. Mitrinović and P. M. Vasić, *History, variations and generalizations of the Čebyšev inequality and the question of some priorities*, Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. Fiz. **461**, 1974, 1–30.
170. B. Mond and J. E. Pečarić, *A Mixed Means Inequality*, Austral. Math. Soc. Gazette **23** (1996), 67–70.
171. P. Montel, *Sur les fonctions convexes et les fonctions sousharmoniques*, Journal de Math. (**9**) **7** (1928), 29–60.
172. R. F. Muirhead, *Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters*, Proc. Edinburgh Math. Soc. **21** (1903), 144–157.
173. J. Nash, *Non-cooperative games*, Ann. of Math. **54** (1951), 286–295.
174. T. Needham, *A visual explanation of Jensen's Inequality*, Amer. Math. Monthly **100** (1993), 768–771.
175. I. Newton, *Arithmetica universalis: sive de compositione et resolutione arithmetica liber*, 1707.
176. C. P. Niculescu, *Convexity according to the geometric mean*, Math. Inequal. Appl. **3** (2000), 155–167.
177. C. P. Niculescu, *A new look at Newton's Inequalities*, J. Inequal. Pure Appl. Math. (JIPAM) **1** (2000), no. 2, article 17.
178. C. P. Niculescu, *A multiplicative mean value and its applications.* In: Inequality Theory and Applications, Vol. 1 (Y. J. Cho, S. S. Dragomir and J. Kim, eds.), Nova Science Publishers, Huntington, New York, 2001, 243–255.
179. C. P. Niculescu, *An extension of Chebyshev's inequality and its connection with Jensen's inequality*, J. Inequal. Appl. **6** (2001), 451–462.
180. C. P. Niculescu, *A note on the Hermite–Hadamard Inequality*, The Mathematical Gazette, July 2001, 48–50.
181. C. P. Niculescu, *Choquet theory for signed measures*, Math. Inequal. Appl. **5** (2002), 479–489.
182. C. P. Niculescu, *The Hermite–Hadamard inequality for functions of several variables*, Math. Inequal. Appl. **5** (2002), 619–623.
183. C. P. Niculescu, *Convexity according to means*, Math. Inequal. Appl. **6** (2003), 571–579.
184. C. P. Niculescu, *An extension of the Mazur–Ulam Theorem.* In: *Global Analysis and Applied Mathematics* (K. Tas, D. Krupka, O. Krupkova and D. Baleanu, eds.), American Institute of Physics Conference Proceedings **729**, New York, 2004, 248–256.
185. C. P. Niculescu, *Interpolating Newton's Inequalities*, Bull. Math. Soc. Sci. Math. Roumanie **47(95)** (2004), 67–83.
186. C. P. Niculescu and L.-E. Persson, *Old and New on the Hermite–Hadamard Inequality*, Real Anal. Exchange, **29** (2003/04), 663–686.
187. C. P. Niculescu and F. Popovici, *A Note on the Denjoy–Bourbaki Theorem*, Real Anal. Exchange, **29** (2003/04), 639–646.
188. C. P. Niculescu and F. Popovici, *A Refinement of Popoviciu's Inequality.* Preprint, 2004.

189. C. P. Niculescu and F. Popovici, *The Extension of Majorization Inequalities within the Framework of Relative Convexity*. Preprint, 2004.
190. A. Nissenzweig, $w^\star$*-sequential convergence*, Israel J. Math. **22** (1975), 266–272.
191. L. Nikolova, L.-E. Persson and T. Zachariades, *On Clarkson's Inequality, Type and Cotype of Edmunds–Triebel Logarithmic Spaces*, Arch. Math. **80** (2003), 165–176.
192. B. Opic and A. Kufner, *Hardy Type Inequalities*, Longman, Harlow, 1990.
193. R. Osserman, *Curvature in the Eighties*, Amer. Math. Monthly **97** (1990), 731–756.
194. A. M. Ostrowski, *Sur quelques applications des fonctions convexes et concaves au sens de I. Schur*, J. Math. Pures Appl. (9) **31** (1952), 253–292.
195. J. E. Pečarić, *A simple proof of the Jensen–Steffensen inequality*, Amer. Math. Monthly **91** (1984), 195–196.
196. J. E. Pečarić, F. Proschan and Y. C. Tong, *Convex functions, partial orderings and statistical applications*, Academic Press, New York, 1992.
197. J. E. Pečarić and V. Šimić, *Stolarski–Tobey mean in n variables*, Math. Inequal. Appl. **2** (1999), 325–341.
198. J. E. Pečarić and K. B. Stolarsky, *Carleman's inequality: history and new generalizations*, Aequationes Math. **61** (2001), 49–62.
199. R. R. Phelps, *Convex Functions, Monotone Operators, and Differentiability*, 2nd edition, Lecture Notes in Math. **1364**, Springer-Verlag, Berlin, 1993.
200. R. R. Phelps, *Lectures on Choquet's Theorem*, 2nd edition, Lecture Notes in Math. **1757**, Springer-Verlag, Berlin, 2001.
201. A. O. Pittenger, *The logarithmic mean in n variables*, Amer. Math. Monthly **92** (1985), 99–104.
202. G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus Analysis*, Vols. I and II, Springer-Verlag, 1925. English edition, Springer-Verlag, 1972.
203. T. Popoviciu, *Sur quelques propriétés des fonctions d'une ou de deux variables réelles*, Mathematica (Cluj) **8** (1934), 1–85.
204. T. Popoviciu, *Notes sur les fonctions convexes d'ordre superieur* (IX), Bull. Math. Soc. Roumaine Sci. **43** (1941), 85–141.
205. T. Popoviciu, *Les Fonctions Convexes*, Hermann, Paris, 1944.
206. T. Popoviciu, *Sur certaines inégalités qui caractérisent les fonctions convexes*, Analele Ştiinţifice Univ. "Al. I. Cuza", Iasi, Secţia Mat. **11** (1965), 155–164.
207. W. Pusz and S. L. Woronowicz, *Functional calculus for sesquilinear forms and the purification map*, Rep. Math. Phys. **8** (1975), 159–170.
208. R. Rado, *An inequality*, J. London Math. Soc. **27** (1952), 1–6.
209. M. M. Rao and Z. D. Ren, *Theory of Orlicz Spaces*, Marcel Dekker, 1991.
210. M. Rădulescu and S. Rădulescu, *Generalization of Dobrusin's Inequalities and Applications*, J. Math. Anal. Appl. **204** (1996), 631–645.

211. M. Renardy and R. C. Rogers, *An Introduction to Partial Differential Equations*, Springer-Verlag, Berlin, 1993.
212. A. W. Roberts and D. E. Varberg, *Convex Functions*, Academic Press, New York and London, 1973.
213. R. T. Rockafellar, *Convex Analysis*, Princeton Mathematical Series **28**, Princeton Univ. Press, Princeton, New Jersey, 1970.
214. G. Rodé, *Eine abstrakte Version des Satzes von Hahn–Banach*, Arch. Math. (Basel) **31** (1978), 474–481.
215. L. J. Rogers, *An extension of a certain theorem in inequalities*, Messenger of Math. **17** (1888), 145–150.
216. P. Roselli and M. Willem, *A convexity inequality*, Amer. Math. Monthly **109** (2002), 64–70.
217. S. Rosset, *Normalized symmetric functions, Newton's inequalities and a new set of stronger inequalities*, Amer. Math. Monthly **96** (1989), 815–819.
218. W. Rudin, *Real and Complex Analysis*, 3rd edition, McGraw–Hill, New York, 1987.
219. S. Saks, *Sur un théorème de M. Montel*, C. R. Acad. Sci. Paris **187** (1928), 276–277.
220. T. Sasaki, Y. Kanada and S. Watanabe, *Calculation of discriminants of higher degree equations*, Tokyo J. Math. **4** (1981), 493–503.
221. R. A. Satnoianu, *Erdős–Mordell–Type Inequalities in a Triangle*, Amer. Math. Monthly **110** (2003), 727–729.
222. I. Schur, *Über die charakteristischen Wurzeln einer linearen Substitution mit einer Anwendung auf die Theorie der Integralgleichungen*, Math. Ann. **66** (1909), 488–510.
223. I. Schur, *Bemerkungen zur Theorie der beschränken Bilinearformen mit unendlich vielen Veränderlichen*, J. Reine Angew. Math. **140** (1911), 1–28.
224. I. Schur, *Über eine Klasse von Mittelbildungen mit Anwendungdie Determinanten*, Sitzungsber. Berl. Math. Ges. **22** (1923), 9–20.
225. W. Sierpiński, *Sur la question de la mesurabilité de la base de M. Hamel*, Fund. Math. **1** (1920), 105–111.
226. W. Sierpiński, *Sur les fonctions convexes mesurables*, Fund. Math. **1** (1920), 125–129.
227. S. Simic, *Proposed Problem 10945*, Amer. Math. Monthly **109** (2002), 475.
228. J. F. Steffensen, *On certain inequalities between mean values, and their application to actuarial problems*, Skand. Aktuarietidskr., 1918, 82–97.
229. J. F. Steffensen, *On certain inequalities and methods of approximation*, J. Inst. Actuaries **51** (1919), 274–297.
230. J. F. Steffensen, *On a generalization of certain inequalities by Tchebycheff and Jensen*, Skand. Aktuarietidskr. **8** (1925), 137–147.
231. J. Stoer and C. Witzgall, *Convexity and Optimization in Finite Dimensions*, Vol. I, Springer-Verlag, Berlin, 1970.
232. K. B. Stolarsky, *Generalizations of the logarithmic mean*, Math. Mag. **48** (1975), 87–92.
233. O. Stolz, *Grunzüge der Differential und Integralrechnung*, Vol. 1, Teubner, Leipzig, 1893.

234. J. Sylvester, *On Newton's Rule for the discovery of imaginary roots of equations.* In: The Collected Mathematical Papers of James Joseph Sylvester, Vol. II (1854–1873), Cambridge Univ. Press, 1908, 493–494.
235. J. Sylvester, *On an elementary proof and generalization of Sir Isaac Newton's hitherto undemonstrated rule for discovery of imaginary roots.* In: The Collected Mathematical Papers of James Joseph Sylvester, Vol. II (1854–1873), Cambridge Univ. Press, 1908, 498–513.
236. M. Tomić, *Théorème de Gauss relatif au centre de gravité et son application*, Bull. Soc. Math. Phys. Serbie **1** (1949), 31–40.
237. Y. L. Tong, *Probability inequalities in multivariate distributions*, Academic Press, New York, 1980.
238. T. Trif, *Convexity of the Gamma Function with Respect to Hölder Means.* In: *Inequality Theory and Applications*, Vol. 3 (Y. J. Cho, J. K. Kim, and S. S. Dragomir eds.), Nova Science Publishers, Inc., New York, 2003, 189-195.
239. J. Väisälä, *A Proof of the Mazur–Ulam Theorem*, Amer. Math. Monthly **110** (2003), 633–635.
240. C. Villani, *Topics in Optimal Transportation*, Graduate Studies in Math. **58**, Amer. Math. Soc., Providence, R.I., 2003.
241. A. Vogt, *Maps which preserve equality of distance*, Studia Math. **XLV** (1973), 43–48.
242. V. Výborný, *The Hadamard three-circles theorems for partial differential equations*, Bull. Amer. Math. Soc. **80** (1974), 81–84.
243. R. Webster, *Convexity*, Oxford Univ. Press, Oxford–New York–Tokyo, 1994.
244. H. Weyl, *Das asymtotische Verteilungsgesetz der Eigenwerte lineare partieller Differentialgleichungen*, Math. Ann. **71** (1912), 441–479.
245. H. Weyl, *Inequalities between two kinds of eigenvalues of a linear transformation*, Proc. Nat. Acad. Sci. USA **35** (1949), 408–411.
246. H. Wielandt, *An extremum property of sums of eigenvalues.* Proc. Amer. Math. Soc. **6** (1955), 106–110.
247. W. H. Young, *On classes of summable functions and their Fourier series*, Proc. Roy. Soc. London, Ser. A **87** (1912), 225–229.
248. A. J. Yudine, *Solution of two problems on the theory of partially ordered spaces*, Dokl. Akad. Nauk SSSR **23** (1939), 418–422.
249. T. Zamfirescu, *The curvature of most convex surfaces vanishes almost everywhere*, Math. Z. **174** (1980), 135–139.
250. T. Zamfirescu, *Nonexistence of curvature in most points of most convex surfaces*, Math. Ann. **252** (1980), 217–219.
251. T. Zamfirescu, *Curvature properties of typical convex surfaces*, Pacific J. Math. **131** (1988), 191–207.
252. W. P. Ziemer, *Weakly Differentiable Functions, Sobolev Spaces and Functions of Bounded Variation*, Graduate Texts in Math. **120**, Springer-Verlag, Berlin, 1989.
253. A. Zygmund, *Trigonometric Series*, Vols. 1 and 2, 2nd edition, Cambridge Univ. Press, 1959.

Index